

KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT TOEGEPASTE WETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK, SCD
Kasteelpark Arenberg 10, 3001 Leuven
in samenwerking met
FACULTEIT GENEESKUNDE
DEPARTEMENT NEUROWETENSCHAPPEN, Lab.Exp.ORL
Kapucijnenvoer 33, 3000 Leuven

ADAPTIVE NOISE REDUCTION ALGORITHMS FOR SPEECH INTELLIGIBILITY IMPROVEMENT IN DUAL MICROPHONE HEARING AIDS

Promotor:
Prof. dr. ir. M. Moonen
Prof. dr. J. Wouters

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

door

Jean-Baptiste MAJ

Juni 2004



KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT TOEGEPASTE WETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK, SCD
Kasteelpark Arenberg 10, 3001 Leuven
in samenwerking met
FACULTEIT GENEESKUNDE
DEPARTEMENT NEUROWETENSCHAPPEN, Lab.Exp.ORL
Kapucijnenvoer 33, 3000 Leuven

ADAPTIVE NOISE REDUCTION ALGORITHMS FOR SPEECH INTELLIGIBILITY IMPROVEMENT IN DUAL MICROPHONE HEARING AIDS

Jury:
Prof. dr. ir. P. Verbaeten, voorzitter
Prof. dr. ir. M. Moonen, promotor
Prof. dr. J. Wouters, promotor
Prof. dr. ir. D. Van Compernolle
Prof. dr. ir. S. Van Huffel
Prof. dr. ir. W.A. Dreschler (AMC, Amsterdam)
Dr. Y. Laprie (CNRS, France)

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

door

Jean-Baptiste MAJ

U.D.C. 534.77

Juni 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/2004/7515/27

ISBN 90-5682-489-9

à ma Mère,

à Gérard,

à Nicole.

Acknowledgements

The research work that I have carried out during the four years of my PhD could not have been accomplished without the support and the collaboration of other people.

Above all, I would like to express my gratitude to Professor Marc Moonen and Professor Jan Wouters. Although I had a limited knowledge of English, they decided to trust me and to give me the opportunity to prepare a PhD in their laboratories. Both of them have followed my research competently and regularly made relevant comments concerning the analysis of my results. They know how to manage their teams and to create a cheerful and fertile environment to carry out research in the best conditions. They also know how to motivate their students when the spirits are low.

Many thanks to the members of my advisory committee, Professor Sabine Van Huffel and Professor Dirk Van Compernolle for their interest in my research, their suggestions and for the constructive comments on my text. I would like to thank Professor Petrus Verbaeten, President of the Jury, Professor Wouter A. Dreschler and Doctor Yves Laprie, the members of the Jury, for taking time to consider my thesis.

I also would like to thank all my colleagues of the Lab. Exp. ORL (Ann, Ann, Astrid, Bram, Heleen, Johan, Koen, Lies, Olivier, Stijn, Tim and Thomas) and also the 'old' colleagues (Els, Jeff, Luc and Guoping) for the nice time that we have had together. We have solved so many problems and performed so many experiments together. I think I will never forget the 'bzi-bzu' experiments of Astrid and Johan, and by far, I preferred the experiments of Els and Heleen where I was allowed to sleep.

Furthermore, I would like to thank the FWO Fund for Scientific Research - Flanders and the IWT for the financial support during my stay in Belgium. Of course, I cannot forget GNReSound (especially Jos, Rob and René) and Cochlear Ltd (especially Bas, Jan, Kyri) for providing me the different behind-the-ear devices and for their contributions to my research. I do not want to

forget the teachers at the CLT Leuven who allowed me to improve my English significantly through all these years.

J'aimerais remercier Gérard et Olivier sans qui, le projet d'effectuer le service militaire au titre de la coopération, et de ce fait, d'effectuer mon doctorat en Belgique, n'aurait aboutit. De plus, votre dévouement envers ma personne m'a toujours beaucoup touché. Nombreux de vos conseils, ajoutés à ceux de Bernard et de Michel, m'ont permis d'aboutir à ce doctorat.

La vie associative prend une place importante dans ma vie. Merci aux incorruptibles du sport du dimanche (souvent de bonne heure, de bonne humeur !) comme le ski club, le CSAR et le 'natuurvrienden', pour tous ces très bons moments passés ensemble ! Merci aussi aux amis dont j'ai fait la connaissance en Belgique et qui m'ont permis de vivre agréablement à Leuven (Auxane, Catherine, Colette, Luc, Marcello, Nicolas et bien sur Rossella), mais sans oublier, Rachèle et Teddy, restés eux en France.

Finalement, je voudrais remercier Véronique et Pascal pour avoir pris soin de 'La Grande' durant mon absence. La dernière mais pas la moindre, ma Mère, à qui je dois tout ce que je suis aujourd'hui 'L'enfant est l'argile, la mère le potier'... merci.

15 Avril 2004

Abstract

Hearing-impaired listeners and hearing aid users have great difficulty understanding speech in a noisy background. This problem has motivated the development and the use of noise reduction algorithms to improve the speech intelligibility in hearing aids. In this thesis, two adaptive noise reduction algorithms are presented for dual microphone behind-the-ear hearing aids and evaluated theoretically, physically and perceptually. The first noise reduction algorithm is optimized to be implemented in the most recent digital hearing aids. This strategy works with assumptions about the position of the speaker and the characteristics of the microphones. A perceptual experiment demonstrates that this strategy gives a larger improvement of the speech understanding than the actual noise reduction algorithms currently implemented in commercial hearing aids. The second noise reduction algorithm works without assumptions about the position of the speaker and the characteristics of the microphones. This noise reduction strategy does not need a calibration procedure unlike the first technique developed in this study. A perceptual evaluation demonstrates that this technique has additional advantages in speech intelligibility in adverse listening conditions relative to the first strategy. The perceptual evaluations are performed with normal hearing listeners and hearing impaired subjects in moderately reverberant environments which are representative of real living rooms.

Korte Inhoud

Slechthorenden en hoorapparaatgebruikers hebben grote moeilijkheden met spraakverstaan in een ruizige omgeving. Dit probleem motiveert de ontwikkeling en het gebruik van ruisonderdrukkingsalgoritmen om spraakverstaan te verbeteren. In dit proefschrift worden twee adaptieve ruisonderdrukkingsalgoritmen aangebracht voor achter-het-oor hoorapparaten met twee microfoons. Ze worden theoretisch, fysisch en perceptueel geëvalueerd. Het eerste ruisonderdrukkingsalgoritme is geoptimaliseerd voor implementatie in de meest recente digitale hoorapparaten. Deze strategie is gebaseerd op veronderstellingen omtrent de positie van de spreker en de karakteristieken van de microfoons. Een perceptueel experiment toont aan dat deze strategie een verbetering betekent voor het spraakverstaan vergeleken met de huidige ruisonderdrukkingsalgoritmen die geïmplementeerd worden in commerciële hoorapparaten. Het tweede ruisonderdrukkingsalgoritme maakt geen gebruik van veronderstellingen wat betreft de positie van de spreker en de karakteristieken van de microfoons. Deze ruisonderdrukkingsstrategie heeft geen nood aan een calibratieprocedure in tegenstelling tot de eerste techniek die tijdens dit onderzoek ontwikkeld werd. Een perceptuele evaluatie demonstreert dat deze techniek een bijkomende winst ople-vert in spraakverstaan relatief ten opzichte van de eerste strategie in ongunstige luisteromstandigheden. De perceptuele evaluatie werd uitgevoerd met normaalhorenden en patienten met gehoorverlies in een matig reverberante omgeving en kan als een typisch voorbeeld worden beschouwd van de effecten in een echte leefomgeving.

Glossary

Mathematical Notation

a	scalar a
a^*	complex conjugation of a
$\mathbf{a}$	vector $\mathbf{a}$
$\mathbf{a}^i$	ith column of a matrix $\mathbf{A}$
$\mathbf{A}$	matrix $\mathbf{A}$
$\mathbf{A}^T$	transpose of a matrix $\mathbf{A}$
$\mathbf{A}^H$	Hermitian transpose of a matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	inverse of a matrix $\mathbf{A}$
$\mathbf{A}^{i,j}$	(i,j)-th element of a matrix $\mathbf{A}$
$r_{x,x}(l)$	autocorrelation function of a stochastic process $x(k)$
$r_{x,y}(l)$	cross-correlation function of stochastic processes $x(k)$ and $y(k)$
$x(k)$	stochastic process
x_k	sample of a time series $\mathbf{x}_k$
$\mathbf{x}_k$	time series of a stochastic process $x(k)$
$\mathbf{x}_{j,k}$	time series of a stochastic process $x(k)$ at the j-th microphone
$H_{x,y}(f)$	power transfer function between $\mathbf{x}_k$ and $\mathbf{y}_k$
$\hat{P}_{x,x}(f)$	estimate of the power spectral density for a vector $\mathbf{x}_k$
$P_{x,x}(f)$	power spectral density of vector $\mathbf{x}_k$
$P_{x,y}(f)$	cross-power spectral density of vectors $\mathbf{x}_k$ and $\mathbf{y}_k$
$P_{x,y}^{speech}(f)$, $P_{x,y}^{noise}(f)$	power spectral density for a speech and a noise signal
$P_{VADerror}^{speech}(f)$, $P_{VADerror}^{noise}(f)$	power spectral density for a speech and a noise signal including VAD errors
$\mathbf{R}_{x,x}$	correlation matrix of a vector $\mathbf{x}_k$

∂	derivative function
$\Gamma_{x,y}(f)$	complex coherence function between a vector $\mathbf{x}_k$ and a vector $\mathbf{y}_k$
$\Gamma_{x,y}^{speech}(f)$, $\Gamma_{x,y}^{noise}(f)$	complex coherence function between a vector $\mathbf{x}_k$ and a vector $\mathbf{y}_k$ for a speech and noise signal
$*$	convolution
$\mu(k)$	mean function of a stochastic process $\mathbf{x}_k$
Im	imaginary part
Re	real part
$\|.\|$	absolute value
$\|.\|_2$	L_2-norm
$\mathcal{E}\{.\}$	expectation operator
$\mathcal{F}\{.\}$	Discrete-Time Fourier Transform operator
$a \approx b$	a is approximately b
$diag\{\mathbf{a}\}$	square diagonal matrix with a vector $\mathbf{a}$ as diagonal
ln	natural logarithm

Fixed Symbols

$\alpha(f)$	signal-to-noise ratio at a microphone signal
β	ratio between an inner delay and an external delay of a hardware directional microphone
γ	constrained value of a white noise gain
$\delta_c(f)$, $\delta_h(f)$	weighted-delay of a software directional microphone
ϵ_s^2, ϵ_n^2	signal distortion and residual noise energy
θ	azimuth coordinate
ϑ	average absorption coefficient in the room
λ_u, λ_n	exponential weighting factors for speech-and-noise and noise-only signal
μ	speech distortion parameter
ν	adaptation constant of a NLMS procedure
$\rho(f)$	sensor-to-environmental noise ratio
σ, η	generalized singular values of matrices $\mathbf{R}_{u,u}$ and $\mathbf{R}_{n,n}$
$\bar{\sigma}$, $\bar{\eta}$	generalized eigenvalues of matrices $\mathbf{R}_{u,u}$ and $\mathbf{R}_{n,n}$
ς	positive constant
τ	internal delay of the software directional microphone
ϕ	elevation coordinate

Δ_1, Δ_2	delay of the first and the second stage of the two-stage adaptive beamformer
$\mathbf{\Sigma}_s + \mathbf{\Sigma}_{n1}$	speech-and-noise subspace
$\mathbf{\Sigma}_{n2}$	noise-only subspace
a	weight of a software directional microphone
b	total number of one third octave bands
c	speed of sound propagation (c=340m/s)
d	constant intermicrophone distance
f	frequency-domain variable
f_s	sampling frequency
g, h	multiplication and power constant of the generalized power spectral subtraction method
$h(k)$	stochastic process of a transfer function between a speaker and a listener
k, k'	discrete-time index
$\Delta gain$, $\Delta phase$	deviations in gain and phase of a microphone
m, n	multiplication and power constant of the generalized frequency response of the Wiener filter
p, q	length of matrices $\mathbf{U}_k$ and $\mathbf{N}_k$
$s(k)$, $n(k)$, $u(k)$	stochastic processes of a speech signal, a noise signal and a noisy speech signal
t	time variable
$w_1^{A2B}(f)$, $w_2^{A2B}(f)$	coefficient of the first and the second filter of the two-stage adaptive beamformer at the frequency f
$w(i,j)$	(i,j)-element of the matrix $\mathbf{W}_{WF,k}$
$\mathbf{d}(f)$	steering vector
$\mathbf{do}_{j,k}$	time series of the j-th microphone of the hearing aid DO
$\mathbf{e}_k$	error of the estimate
$\mathbf{e}_{s,k}$, $\mathbf{e}_{n,k}$	signal distortion and residual noise
$\mathbf{h}_k$	acoustic impulse response between a source and a microphone
$\mathbf{n}_k$, $\mathbf{s}_k$	stacked noise data vector and speech data vector
$\mathbf{n}_{j,k}$	stacked noise data vector of the j-th microphone
$\mathbf{n}_{\mathbf{ref},k}$	time series of the noise reference signal
$\hat{\mathbf{s}}_k$	estimate of the stacked speech data vector by a linear filter
$\mathbf{u}_k$	stacked speech-and-noise data vector
$\mathbf{u}_{j,k}$	stacked speech-and-noise data vector of the j-th microphone
$\mathbf{w}(f)$	stacked filter vector of fixed beamformers
$\mathbf{w}_j^{A2B}$	stacked j-th filter vector of the two-stage adaptive beamformer
$\mathbf{w}_j^{WF}$, $\mathbf{w}_{j,k}^{WF}$	stacked filter vector of the j-th microphone of the optimal filtering technique

$\mathbf{w}_{WF,k}$	stacked filter vector of the optimal filtering technique
A_i	weights for the audibility function of the speech intelligibility at ith one-third octave bands
Det	determinant of a matrix
$D(\theta)$	spatial characteristic of the directional microphone
D_F	derivation of the factor F
$DI(f)$	directivity index
DI_{AI}	intelligibility-weighted directivity index
DO	hearing aid with one hardware directional microphone and one omnidirectional microphone
DO_{Hd}, DO_{Fo}	hardware directional microphone and omnidirectional microphone of the hearing aid DO
F_{SN}, F_N	speech-and-noise and noise-only frames
$F_{SN}^{VADerror}$, $F_N^{VADerror}$	speech-and-noise and noise-only frames including VAD errors
G_{AI}	gain in intelligibility weighted SNR based on the Articulation Index
$G(f)$	array gain
G_{SII}	gain in intelligibility weighted SNR based on the Speech Intelligibility Index
$G_w(f)$	white noise gain
I_i	weights for the importance of the band for the speech intelligibility at ith one-third octave bands
L	length of a time series
L_1^{A2B}, L_2^{A2B}	size of the first and second filter of the two-stage adaptive beamformer
L^{WF}	size of the filter per channel of the optimal filtering technique
M	number of microphones
OO	hearing aid with two omnidirectional microphones
OO_{Fo}, OO_{Ro}	front and rear omnidirectional microphone of the hearing aid OO
OO_{Sd}	software directional microphone of the hearing aid OO
$OO_{Ro}^{Dev}(f)$	rear omnidirectional microphone of the hearing aid OO including a gain and a phase deviation at the frequency f
$N_{Perfect}$	samples which belong to the noise-only period with a 'perfect' VAD
$N_{RealTime}$	samples which are correctly classified as noise-only period by the VAD
Per	percentage correctly detected by the VAD
R	rank of the speech clean model
S	surface area of the room
$SN_{Perfect}$	samples which belong to the speech-and-noise period with a 'perfect' VAD

$SN_{RealTime}$	samples which are correctly classified as speech-and-noise period by the VAD
SNR_{AI}	intelligibility-weighted SNR based on the Articulation Index
SNR_{SII}	intelligibility-weighted SNR based on the Speech Intelligibility Index
T_n, T_D, T_w	threshold values
T_{60}	reverberation time of a room
$W_{WF}(f)$	generalized frequency response of the Wiener filter
V	volume of the room
VAD_{error}	model of VAD errors
$\mathbf{B}$	Griffith-Jim blocking matrix
$\mathbf{I}$	identity matrix
$\mathbf{J}_{MSE}$	MSE cost function
$\mathbf{N}_k$, $\mathbf{U}_k$	noise-only data matrix and speech-and-noise data matrix
$\mathbf{Q}_k$	matrix containing generalized singular vectors of $\mathbf{U}_k$ and $\mathbf{N}_k$
$\mathbf{R}_{U,k}$, $\mathbf{R}_{N,k}$	upper triangular speech-and-noise and noise-only data matrices
$\mathbf{W}_k$	linear filter matrix
$\mathbf{W}_{WF}$, $\mathbf{W}_{WF,k}$	Wiener filter matrix
$\mathbf{X}$	matrix containing generalized singular vectors of $\mathbf{U}_k$ and $\mathbf{N}_k$
$\bar{\mathbf{X}}$	matrix containing generalized eigenvalue vectors of $\mathbf{R}_{u,u}$
$\mathbf{P}_{N,k}$, $\mathbf{P}_{U,k}$	orthogonal matrices
$\mathbf{Y}$, $\mathbf{Y}_k$	matrix containing generalized singular vectors of $\mathbf{U}_k$
$\mathbf{Z}$, $\mathbf{Z}_k$	matrix containing generalized singular vectors of $\mathbf{N}_k$

Acronyms and Abbreviations

A2B	Two-Stage Adaptive Beamformer
ADC	Analog Digital Converter
AEC	Acoustics Echo Canceller
AI	Articulation Index
AIC	Akaike Information Criterion
ANC	Adaptive Noise Canceller
ANOVA	Analysis Of Variance
BSS	Blind Source Separation
BTE	Behind-The-Ear

CCF	Complex Coherence Function
C-PSD	Cross-Power Spectral Density
DAC	Digital Analog Converter
DI	Directivity Index
DTFT	Discrete-Time Fourier Transform
dB	decibel
DSP	Digital Signal Processor
ECM	Error Covariance Matrix
EEG	Electroencephalographic
e.g.	*exempli gratia*: for example
EOG	Electrooculographic
EVD	Eigenvalue Decomposition
ERP	Event-Related Potentials
GSC	Generalized Sidelobe Canceller
GSVD	Generalized Singular Value Decomposition
ICA	Independent Component Analysis
ICRA	International Collegium of Rehabilitative Audiology
i.e.	*id est*: that is
ITC	In-The-Canal
ITE	In-The-Ear
KLT	Karhunen-Loève Transform
LMS	Least Mean Square
LR	Likelihood Ratio
LS	Least Square
LTASS	Long-Term Average Speech Spectrum
MAPD	Modified Probability Distribution Function
MMSE	Minimum Mean Square Error
MSE	Mean Square Error
MDL	Minimum Description Length
PCA	Principal Component Analysis
PSD	Power Spectral Density
PTA	Pure Tone Average
PTF	Power Transfer Function
SII	Speech Intelligibility Index
SNR	Signal-to-Noise Ratio
SRT	Speech Reception Threshold
SVD	Singular Value Decomposition
VAD	Voice Activity Detector

Contents

Abstract iii

Korte Inhoud v

Glossary vii

1 Introduction 1

1.1 Speech intelligibility 1

1.1.1 Speech signal 2

1.1.2 Noise signal 2

1.2 Hearing impairment and hearing aids 4

1.3 Main issues in digital hearing aids 5

1.3.1 Feedback cancellation 5

1.3.2 Noise reduction 7

1.3.3 Amplification and compression 7

1.4 Outline of the thesis and main contributions 8

2 Characterization of Signals and Noise Reduction Algorithms 13

2.1 Recorded signals at the microphones 14

2.1.1 Acoustic environment 14

2.1.2 Microphone characteristics 15

2.2 Noise reduction algorithms 16
2.2.1 Performance metrics 16
2.2.2 Voice activity detector 19
2.2.3 State-of-the-art noise reduction algorithms 22
2.3 Conclusion 34

3 Two-stage adaptive beamformer **37**
3.1 Two-stage adaptive beamformer 38
3.2 Theoretical Analysis 41
3.2.1 Derivation of the equations 41
3.2.2 Results and discussion 43
3.3 Optimization of the algorithm 44
3.3.1 Implementation of the A2B 45
3.3.2 Physical evaluation 47
3.3.3 Perceptual evaluations 50
3.3.4 Discussion 54
3.4 Large scale evaluation 55
3.4.1 Experimental set-up 55
3.4.2 Physical evaluations 58
3.4.3 Perceptual evaluations 58
3.4.4 Test materials 60
3.4.5 Results and discussion 61
3.4.6 Summary 71
3.5 Comparison of the A2B with an adaptive directional microphone 72
3.5.1 Noise reduction algorithms 73
3.5.2 Experimental set-up 74
3.5.3 Physical evaluation 76

3.5.4 Perceptual evaluation 78
3.5.5 Results and discussions 78
3.6 Conclusion 84

4 SVD-based optimal filtering technique 87
4.1 Optimal filtering technique 88
4.2 Theoretical evaluation 91
4.2.1 Derivation of the equations 91
4.2.2 Experimental set-up 92
4.2.3 Results 93
4.3 SVD-based optimal filtering technique 97
4.4 Physical evaluation 98
4.4.1 Experimental set-up 99
4.4.2 Results and discussion 103
4.5 Real time implementation 110
4.5.1 Speech distortion parameter 110
4.5.2 Real time implementation 110
4.6 Improvement of the VAD performance 116
4.6.1 Method 116
4.6.2 Experimental set-up 117
4.6.3 Hearing aids 117
4.6.4 Results 118
4.7 Perceptual evaluations 123
4.7.1 Method 123
4.7.2 Results 126
4.8 Conclusion 132

5 Conclusion **135**

5.1 Two-stage adaptive beamformer 135

5.2 SVD-based optimal filtering technique 138

5.3 Suggestions for further research 140

5.3.1 Two-stage adaptive beamformer 140

5.3.2 Optimal filtering technique 140

Bibliography **143**

Appendices **157**

A Signal processing basics 157

B Linear algebra and matrix decomposition definitions 160

List of Publications **163**

Curriculum Vitae **167**

Chapter 1

Introduction

Most of the time, speech communication takes place in the presence of ambient noise like traffic sounds or speech from competing speakers. In the transfer of speech from speaker to listener, part of the signal can be masked by the ambient noise and speech information can be lost.

In the *first section* of this introductory chapter, the characteristics of the speech and the noise signals are described and the important properties of speech intelligibility are examined. In the *second section*, the effect of the acoustical noise on the speech understanding of the hearing impaired listeners is discussed. Furthermore, the characteristics of hearing aids, which are used to compensate for the hearing loss of hearing impaired persons, are examined. In the *third section*, the functions of digital hearing aids are described, and in the *fourth section*, the aim and the outline of the thesis are discussed.

1.1 Speech intelligibility

One of the most important considerations for speech intelligibility is the *Signal-to-Noise Ratio* (SNR), which is the ratio between the intensity of the speech and the intensity of the noise expressed in decibels (dB).

For instance, in classroom environments, the SNR ranges from +5dB to -7dB (Tyler and Schum, 1995), and in cocktail parties the average SNR ranges from +1dB to -2dB (Plomp, 1978). The required SNR at which 50% of the speech is correctly understood by the listener is called the Speech Reception Threshold (SRT). In this section, the characteristics of the speech and the noise signals are described and the important properties of speech intelligibility are

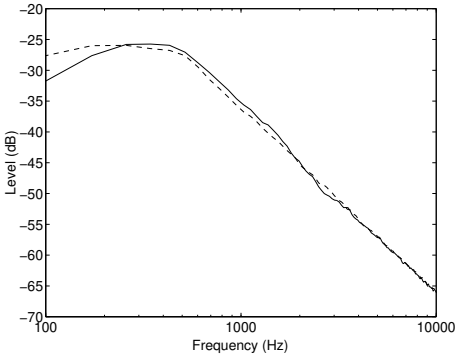


Figure 1.1: Long-Term Average Speech Spectrum (LTASS) for male (−−) and female (−) speakers.

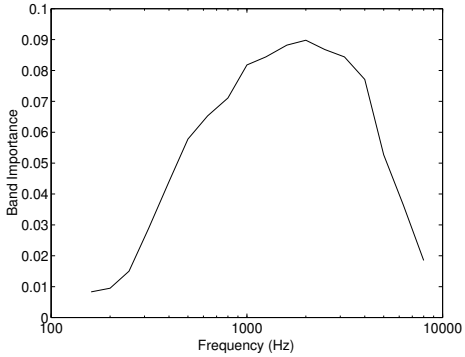


Figure 1.2: Band importance of the frequencies for the speech intelligibility.

examined.

1.1.1 Speech signal

The speech signal is a non-stationary broadband signal. The frequency components range from 100Hz to 10kHz, and the time envelope and the spectrum change continuously. The Long-Term Average Speech Spectrum (LTASS) was studied for twelve languages and an 'universal' LTASS was suggested (Byrne et al., 1994). The LTASS for male and female voices are presented in **figure 1.1** for a normal vocal effort. Most of the speech energy is present at the low frequencies (150Hz-600Hz) with a maximum at about 200Hz for the male voice and 300Hz for the female voice. The main difference between the male and the female LTASS is below 250Hz where the male LTASS exceeds the female LTASS. Otherwise, there are no notable differences between both spectra, which decay by about 6dB/octave above 600Hz.

For speech understanding, the frequencies from about 500Hz to 5kHz carry the most important information. **Figure 1.2** shows the band importance function, representing the contribution to intelligibility as a function of frequency (ANSI, 1997). Speech is also an intermittent signal, where silences occur between the words. Typically in conversation, more than 50% of the time of the speech signal consists of pauses.

1.1.2 Noise signal

All the sounds which are not considered as a desired (speech) signal, can be classified as noise signals. Even speech signals, such as multitalker

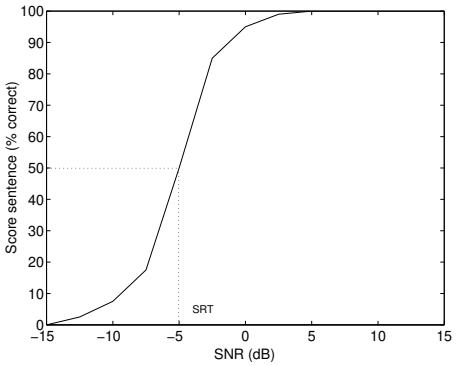


Figure 1.3: Sentence score as a function of the SNR for normal hearing listeners.

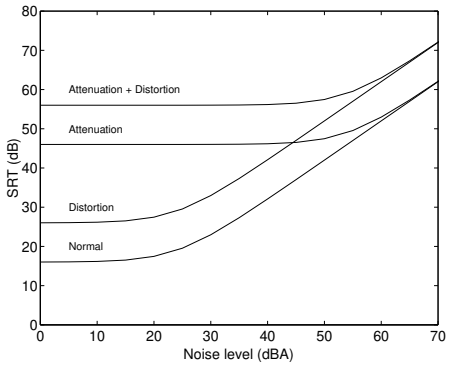


Figure 1.4: Speech level required for 50% sentence score as a function of noise level. The lower curve represents normal hearing. The other curves represent the effect of attenuation, distortion and a combination.

babble or a single concurrent speaker, can be classified as background noise. Spectral and temporal information characterize the noise signals. They can be stationary (e.g. fans) or non-stationary (e.g. traffic), narrowband (e.g. siren) or broadband (e.g. music), and can originate from a localized position or from all directions (diffuse noise).

In natural speech, the information relevant for intelligibility is contained in a 30dB range, symmetric around the average speech level. When for all frequencies the average speech level is 15dB above the level of a masking noise, all information is available. When the average speech signal is 15dB below the noise level, all speech information is masked by the noise. Normal hearing listeners can have an acceptable level of speech intelligibility even when a large part of the speech is masked because the speech has a high level of redundancy. **Figure 1.3** shows the transfer function for Dutch sentences in stationary noise for normal hearing listeners, which is measured by determining the intelligibility score for sentences as a function of the SNR for monaural[1] listening (Plomp, 1978). The shape of the noise spectrum corresponds to the long term average speech spectrum of the sentences. The SRT in noise is approximately -5dB and close to this threshold a small increase in SNR can correspond to an important increase in sentence score by the listener. An improvement of 1dB in SNR can concur to an increase of speech understanding of about 15% in every day speech communication (Plomp and Mimpen, 1979).

[1] A situation in which sounds are presented to one ear only.

1.2 Hearing impairment and hearing aids

Plomp (1978) proposed a model of the SRT as a function of the noise level for hearing impaired persons. He argued that a sensorineural hearing loss does not simply result in an attenuation of all sounds entering the ear, but also in a distortion of these sounds. The effect of the attenuation, the distortion and the combination of both on the SRT as a function of noise is depicted in **figure 1.4**. Compared to the SRT of normal hearing listeners, the attenuation component of the hearing loss influences the SRT at low noise levels but not at high noise levels. With the distortion component, independently of the noise levels, hearing impaired listeners need a higher SNR than normal hearing persons. To overcome the deficits associated with hearing loss, hearing aids are generally used.

There are two basic functions of a hearing aid. The first, the amplification to lower the higher hearing thresholds of the hearing impaired. The second, the compression of the acoustical signal. The latter is necessary to limit the maximum output level (compression limiter) and/or to take into account the reduced dynamic range of the impaired ear (dynamic range compression). The relative uncomfortable loudness levels of a lot of sensorineural hearing impaired listeners are pivotal for this. Besides these basic functions of a hearing aid, additional or supplementary functions are present in modern devices such as *noise reduction* or *feedback cancellation* procedures.

Hearing aids can be categorized into three different types. The first is the *behind-the-ear* (BTE) hearing aid. The microphones, the electronics and the receiver are mounted in a characteristic banana-shaped case. The second is the *in-the-ear* (ITE) hearing aid, which occupies the entire concha (deep center portion of the visible part of the ear) as well as about half of the length of the ear canal. The third type occupies a small portion of the external auditory canal of the ear and is referred to as an *in-the-canal* (ITC) hearing aid. Another distinction between hearing aids can be made on the technology, namely analog or digital.

The first fully digital BTE hearing aid was only introduced in 1995 (Widex Senso) and the market share of digital devices is ever increasing. The differences between analog and digital hearing aids can be made clear using schematic representations (**figure 1.5**). Whereas any hearing aid consists of a microphone system and a receiver part as in- and output of the device, the main differences are present in the functionalities. As such, an analog hearing aid has only a few controls, such as potentiometers or switches. In digital hearing aids, the input signal is converted from the analog to the digital domain (with an Analog-to-Digital Converter: ADC), then processed digitally, with a Digital Signal Processor (DSP), and finally converted to the analog domain (with a Digital-to-Analog Converter: DAC). Digital aids allow remote control

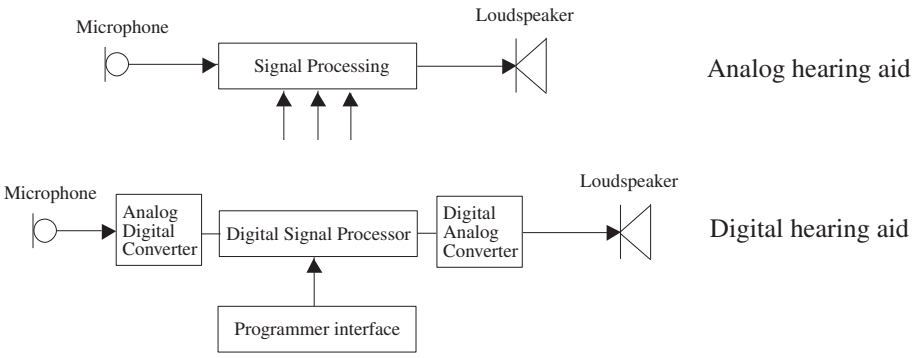


Figure 1.5: Block diagrams of an analog and a digital hearing aid.

and multiple listening programs. Multiple programs may be needed because different acoustical environments may necessitate different hearing aid settings for optimal listening. Furthermore, digital hearing aids allow flexible fitting (via software), more complex processing, several processing schemes and may allow decision making routines in sensing the acoustical environment. Although the size and power consumption of digital circuits is reducing at a faster rate than it is for analog circuits, digital hearing aids are considerably more expensive than analog hearing aids. The latter is, at present, a major disadvantage of digital hearing aids.

1.3 Main issues in digital hearing aids

This thesis is focused on digital BTE hearing aid using two microphones and a general diagram of a device is illustrated in **figure 1.6**. The next paragraphs describe the functions *feedback cancellation*, *noise reduction* and *amplitude-frequency response*.

1.3.1 Feedback cancellation

Feedback oscillation is a major problem with hearing aids. When the patient has an important hearing loss, more gain with the hearing aid is needed. When a large gain is applied, the output of the hearing aid can circulate back to the input of the aid and introduce feedback oscillations (the hearing aid whistles).

A technique for feedback cancellation is to decrease the gain in the fre-

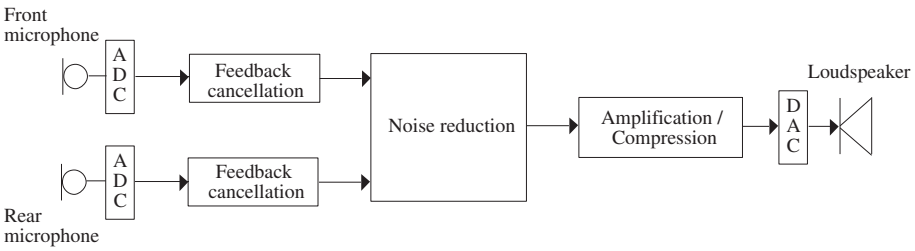


Figure 1.6: General block diagram of a modern digital BTE hearing aid with two microphones.

quency region of the acoustic feedback or to manipulate the phase response of the signal by an all-pass filter. A more complex method for feedback suppression involves adaptive filters which are used to generate an estimate of the feedback signal and subtract it from the feedback signal measured by the microphone, thus reducing its effect. The techniques based on adaptive filters can be classified into two classes: continuously adapting algorithms and non-continuously adapting algorithms. Continuously adapting algorithms minimize the difference between the hearing aid input and a delayed filtered version of the hearing aid output (Bustamante et al., 1989; Joson et al., 1993). However to avoid unstable behaviour, constraints are required on the magnitude of the adaptive filter weights (Kates, 1999) or an additional delay needs to be placed in the forward path of the hearing aid instead of the cancellation path (Siqueira and Alwan, 2000). Noncontinuously adapting algorithms use a detector to determine when strong oscillation occurs. When this condition is met, the forward path of the hearing aid is broken and a broadband probe noise is presented to the receiver of the device. To estimate the feedback path, the adaptive filter is only adapted when the probe signal is active (Kates, 1991; Maxwell and Zurek, 1995). Unfortunately, the sound quality is degraded by injecting noise and by interrupting the audio output. Furthermore, the performance of such techniques may be reduced by the presence of ambient noise or speech at the microphone input.

Feedback cancellation algorithms can still be improved. With classical continuously adapting algorithms, a biased estimate of the feedback path is performed. A recent study presents an unbiased method but only preliminary results are available (Spriet et al., 2002a). Furthermore, the performance of the algorithms is sensitive to the variability of acoustic environments and feedback paths (Kates, 2000; Rafaely et al., 2000).

1.3.2 Noise reduction

Hearing-impaired listeners and hearing aid users have great difficulty understanding speech when there is a noisy background (Duquesnoy and Plomp, 1983; Plomp, 1986; Cox and Alexander, 1991). This problem has motivated the development and the use of noise reduction algorithms to improve the speech intelligibility in hearing aids.

Based on a *single microphone* input the separation of desired and interfering sounds is impossible when the spectral and temporal characteristics are similar. By using *two microphones* spatial information can be used. The latter approach has already been applied for a long time. In 1995, Phonak introduced a directional microphone where the outputs of two omnidirectional microphones were combined (Bachler and Vonlanthen, 1995). In this way, the user could switch between fixed patterns of directionality. Interfering sound sources from the sides or from the back hemisphere can successfully be attenuated in critical speech reception scenarios. More recently, adaptive signal processing has been added to two microphone configurations. In principle, this allows the spatially-selective suppression of jammer sources, even with changing positions. The first adaptive systems have been implemented in commercial digital hearing aids in 2000 (Phonak Claro, GNReSound Canta).

In modern digital hearing aids, the DSPs are becoming smaller and faster. This allows for the implementation of more complex noise reduction strategies. There are two important issues for the noise reduction techniques. A first is the robustness of the algorithms in complex noise scenario, such as multiple jammer sound scenarios or diffuse noise, and a second is the robustness of the algorithms against the variations of the characteristics of the microphones.

1.3.3 Amplification and compression

The fitting of a hearing aid is mainly based on audiometric measures of hearing thresholds and of uncomfortable loudness levels at a set of frequencies. These parameters determine the amplification (the audibility of the sounds) and the compression to be applied in the hearing device. The amplification and the compression systems are the basic functions of hearing aids. Thus, the dynamic range of signals in the environment are increased or decreased so that all signals of interest can fit within the restricted range of a hearing-impaired person.

However, in recent years auditory research has pointed to some specifications related to individual hearing impairments, which are important in hearing aid fitting. Noordhoek et al. (2000a, 2001) have recently carried out extensive psychophysical tests with large numbers of sensorineural hearing im-

paired subjects. The psychophysical measurements were meant to quantify the spectral and temporal resolutions or abilities of the subjects (Noordhoek, 2000). The data show that about 50% of the tested hearing impaired subjects have no suprathreshold deficit. This means that restoring the audibility of sounds in these subjects is most important. The other 50% of the subjects do also have suprathreshold deficits in addition to the audibility-issue: about 30% have mainly spectral deficiencies and about 20% have significant temporal resolution deficits. Moore (Moore and Alkantara, 2001; Baer et al., 2002) investigated methods to quantify so-called intracochlear dead regions along the tonotopical axis.

Results of these studies suggest more individually adjusted hearing aid fittings. These more-than-just-audibility deficits yet to be seriously addressed in hearing aids, but software flexibility of future digital hearing aids may allow taking into account these suprathreshold aspects in the adaptation of the signal processing to the needs of the hearing impaired individual.

1.4 Outline of the thesis and main contributions

This dissertation is focused on the improvement of speech understanding in a noisy environment, using dual microphone digital BTE hearing aids. Two noise reduction algorithms, namely *adaptive beamforming* and *optimal filtering* (Vanden Berghe and Wouters, 1998; Doclo and Moonen, 2002), are optimized, implemented and evaluated in the present study. The assessment of both strategies is performed theoretically, physically and perceptually. In the latter, the evaluation is carried out by fifty-two subjects in highly controlled conditions. Apart from this introductory chapter, this dissertation contains four chapters.

Chapter 2 discusses the properties of the acoustic environment and the characteristics of the microphones. The performance measures considered in this thesis are also presented. These measures predict or estimate the performance of noise reduction algorithms. Since the noise reduction techniques presented in this study rely on a Voice Activity Detector (VAD), an overview of different VAD algorithms is given. Finally, the state-of-the-art of noise reduction strategies is discussed.

Chapter 3 presents the optimization and evaluation of the **adaptive beamformer** technique. In a previous study, the beamformer strategy was evaluated physically and perceptually (Vanden Berghe and Wouters, 1998). In that study, the algorithm was implemented in real time on a PC platform (large computing power) and only a few physical and perceptual tests were carried out.

The aim of the present study is to increase speech intelligibility in current digital hearing aids. In order to accomplish this goal, the robustness of the adaptive beamformer is improved and the strategy is optimized for implementation in modern devices. Extensive physical and perceptual evaluations are carried out to study the influence of jammer noise sound scenarios, spectro-temporal characters of the jammer sounds, hearing impairment and microphone configurations on the performance of the adaptive beamformer. This chapter contains four main studies.

In the *first study*, a theoretical evaluation shows the influence of the distance between the two microphones on hearing aid and the internal noise of the sensors on the performance of the noise reduction strategy. In the *second study*, the adaptive beamformer is optimized for implementation on a wearable research platform, which has a similar computing power as current digital hearing aids. Physical and perceptual evaluations reveal that a significant improvement in speech intelligibility can be obtained relative to the study of Vanden Berghe and Wouters (1998). In the *third study*, a large scale evaluation of the adaptive beamformer is performed. The strategy is evaluated for several speech and noise combinations, several jammer sound source scenarios and two different microphone configurations in hearing aids. In total, 1368 SRT-measurements are carried out. On the one hand, the physical and the perceptual evaluations show that the microphone configuration of the hearing aid has an impact on the performance of the noise reduction strategy. These investigations also demonstrate that the speech intelligibility improvement depends on the jammer sound scenarios. On the other hand, the noise reduction performance is seen to be independent of the spectro-temporal characters of the jammer sound sources and the hearing impairment of the subjects. In the *fourth study*, the adaptive beamformer technique is implemented in a commercial hearing aid and compared to an adaptive directional microphone, similar to the state-of-the-art algorithm in modern commercial digital hearing aids, such as the Phonak Claro and GNReSound Canta. The physical evaluation reveals the importance of an individualized calibration procedure for fixed processing systems, such as the first filter of the adaptive beamformer, to take into account the characteristics of the microphones. The perceptual measurements demonstrates that the adaptive beamformer always performs best in single noise source scenarios. In a multiple jammer sound source scenario, a similar SRT-improvement is obtained with both strategies. The publications related to this chapter are (Maj et al., 2001a; Maj et al., 2001b; Maj et al., 2002b; Wouters et al, 2002; Maj et al., 2003; Maj et al., 2003a; Maj et al., 2003b).

Chapter 4 presents the optimization and the evaluation of the **optimal filtering** technique for dual microphone hearing aid applications (Doclo and Moonen, 2002). In that study, a physical evaluation showed that the optimal filtering technique could give an additional gain in speech intelligibility relative to the adaptive beamformer in highly reverberant conditions and/or when

the assumptions of the latter technique were violated. However, the physical evaluation was performed in unrealistic conditions. The reverberant speech signal was created by using speech material recorded in anechoic conditions and filtered by a simulated acoustic room impulse response. To create the noisy speech signal, an uncorrelated white noise signal was added to the speech signal.

The purpose of the present study is to demonstrate that the optimal filtering technique outperforms the adaptive beamformer in real acoustic conditions. In order to show this, the first physical evaluation of the optimal filtering technique is carried out using acoustic signals recorded in reverberant conditions. This assessment is performed with stationary speech-weighted noise, which means that the spectrum of the speech and the noise signals are roughly the same. Next, the optimal filtering technique is implemented in real time on a PC platform. Using the real time implementation, the first perceptual measurements of the optimal filtering strategy are performed. Since noise reduction strategies are sensitive to the errors of the VAD, a method to improve the performance of the latter is proposed. This chapter contains four main studies.

In the *first study*, the theoretical evaluation demonstrates that the deviations of the microphone characteristics and the VAD errors have an important impact on the noise reduction performance of the strategies. However, the optimal filtering technique is seen to be more robust than the adaptive beamformer against the gain deviation of the microphones. In the *second study*, the physical evaluation of the optimal filtering strategy, computed by a *Generalized Singular Value Decomposition* (GSVD) technique, is performed. The parameters of the latter strategy are optimized by maximizing the gain of speech intelligibility in noise. The physical investigation is carried out in different noise scenarios. In addition, the influence of the VAD errors on the noise reduction performance of both noise reduction techniques is examined. It is shown that the SVD-based optimal filtering technique performs better than the adaptive beamformer in highly reverberant conditions (diffuse noise). This evaluation also confirms that both noise reduction strategies are sensitive to VAD errors. Then, the SVD-based optimal filtering technique is implemented in real time, using a recursive Jacobi-type GSVD-updating strategy. In the *third study*, a method to improve the performance of a VAD is proposed. The physical evaluation of this method shows that the VAD performs well even at low SNRs. In the *fourth study*, the perceptual evaluation of the optimal filtering technique is performed in a single and multiple noise source scenario for two different microphone configurations. Comparing these two noise reduction techniques demonstrates that the SVD-based optimal filtering technique outperforms the adaptive beamformer in adverse listening conditions. The publication related to this chapter are (Maj et al., 2002a; Maj et al., 2002b; Maj et al., 2002c; Maj et al., 2003c; Maj et al., 2004).

Chapter 5 gives an overall conclusion as well as suggestions for future research and experiments.

Chapter 2

Characterization of Signals and Noise Reduction Algorithms

In hearing aid applications, the signals recorded at a microphone are assumed to be equal to $u(k) = h(k) * s(k) + n(k)$, where $s(k)$, $n(k)$ and $h(k)$ are discrete time samples[1] of the speech signal, the noise signal and the transfer function between the speaker and the microphone, respectively. To extract the speech part $s(k)$ from $u(k)$, noise reduction algorithms (interest of this thesis) are used to remove the noise part $n(k)$ and adaptive echo canceller (AEC) or speech dereverberation algorithms are used to remove the acoustic impulse response $h(k)$ (Miyoshi and Kaneda, 1988; Affes and Grenier, 1997; Gannot and Moonen, 2001).

This chapter contains two sections. The *first section* presents the properties of the signals recorded at the microphones of the hearing aids. The characteristics of the acoustic impulse response $h(k)$ and the microphones are discussed. In the *second section*, different noise reduction strategies are examined and their contributions to the improvement of the speech intelligibility in noise is discussed. This section contains three subsections. The first mentions the performance measures of the noise reduction strategies considered in this thesis. Since most of the noise reduction strategies rely on a Voice Activity Detector (VAD), an overview of different VAD algorithms is given in the second subsection. Finally, the third subsection examines the state-of-the-art noise reduction techniques.

[1] In appendix A, a brief review of discrete-time signal processing is presented.

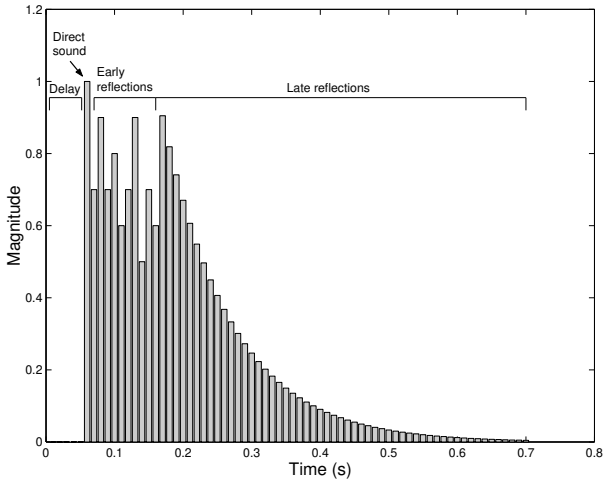


Figure 2.1: Representation of the reverberant sound.

2.1 Recorded signals at the microphones

As mentioned above, the signals recorded at the microphones are function of the acoustic impulse response, the speech and the noise signals. The speech and the environmental noises were described in the first chapter (section 1.1), and in this section, the properties of the acoustic impulse response and the microphones are mentioned.

2.1.1 Acoustic environment

The acoustic impulse response $h(k)$ involves the reverberations of the acoustic environment between the speaker and the microphone of the hearing aids. The linear filter $h(k)$ is defined by three main aspects: the direct sound, the early reflections and the late reflections (**figure 2.1**).

• The component of the sound which travels by the shortest path between the sound source and the listener is called the direct sound. Because the sound needs time to travel from one point to another point, a delay is introduced between the sound source and the listener. The direct sound is important because the information of the signal is carried in an uncontaminated form. For good speech intelligibility, a high level of direct sound is required.

• Somewhat later than the direct sound, the listener catches the sounds which have been reflected off one or more surfaces. These sounds are called the early reflections and are separated from the direct sound in both time and

direction. If any of these reflections are severely delayed (longer than 30ms), they are perceived as echoes and can reduce the speech intelligibility.

• The late reflections or reverberations are sounds which have been reflected many times and are arriving at the listener from all directions at an even later time. Due to the fact that some of the sound is absorbed at each reflection the reverberation eventually dies away. The time that it takes for the sound to die away is called the reverberation time. It is dependent on both the size of the room and the amount of sound absorbed with each reflection.

To characterize the reverberation properties of a room, two measures are often used: the reverberation time, known as T_{60}, and the critical distance. The T_{60} corresponds to the time (in s) for a decrease in sound energy of 60dB. If different parameters of the room such as the volume V (in m^3), the average absorption coefficient ϑ for the surface area S are known, the T_{60} can be calculated by the Norris-Eyring reverberation formula:

$$T_{60} = \frac{-0.161\ V}{S\ ln(1-\vartheta)} \tag{2.1}$$

The critical distance is the distance where the reverberation level equals the direct sound level. The ratio between the direct sound and the reverberant sounds depends on the position of the listener relative to the speaker. If the distance between the listener and the speaker is shorter than the critical distance, the direct sound is predominant. If the distance between the listener and the speaker is greater than the critical distance, the reverberant field is predominant and may affect the intelligibility of the speech.

2.1.2 Microphone characteristics

The noise signal $n(k)$ which is present in the electrical outputs of a microphone originates as an acoustical noise in the environment (section 1.1.2) or as a noise generated within the microphone. In any type of sensor, there are three main noise mechanisms: *thermal noise*, *shot noise* and *flicker noise* (Thompson et al., 2002). Shot noise originates from gate leakage current and flicker noise results from fluctuations in the resistance of circuit elements, which is related to the level of DC current in the device. However, the dominant noise mechanism, in miniature microphones, is thermal noise at all frequencies between 20Hz and 10kHz. It is important that the internal noise of the sensors is low, especially in a microphone array. This can limit the performance of noise reduction techniques and in some strategies, which are sensitive to internal noise, can even amplify this noise.

Most of the signal processing techniques assume that the characteristics of the microphones are identical in the spatial and the spectral domain. This means that all microphones have the same directional pattern and the same

frequency response, e.g. flat and equal to one. However, this is not the case with real sensors. The electrical signal (magnitude and phase) at the output of the microphone changes from sensor to sensor for the same stimulus. Calibration procedures exist to prevent these differences but calibration is not a simple task and the procedure has to be performed for every individual sensor or hearing aid (Sydow, 1994; Buck, 2002; Maj et al., 2003; Maj et al., 2003b). Moreover, the characteristics of the microphone can still drift over the time due to aging and dirt in the acoustic pathway (Tchorz, 2001). Although the effect of aging appears to be limited, dirt in the acoustic pathway can indeed have a strong impact on the characteristics of the microphones.

2.2 Noise reduction algorithms

In the present section, the state-of-the-art of the noise reduction algorithms for improving the speech intelligibility in a noisy background is examined. The first subsection presents the performance measures considered in this dissertation and the second subsection gives an overview of VAD algorithms. Finally, the third section gives a review of the noise reduction techniques.

2.2.1 Performance metrics

In this subsection, the performance measures considered in this thesis are discussed. These measures are mostly based on the average intelligibility gain between the input signal at a microphone of the hearing aid and the output signal of the noise reduction strategy. A preliminary assessment of the noise reduction algorithms is performed with the theoretical evaluation and the physical evaluation should predict the improvement of the speech intelligibility in noise obtained with the perceptual assessment.

2.2.1.1 Theoretical evaluation

The performance measures for the theoretical evaluations are based on the articulation index (AI) (ANSI, 1969). Peterson (1989) and Greenberg et al. (1993) used the AI procedure to develop an intelligibility weighted SNR:

$$SNR_{AI} = \sum_{i=1}^{b} I_i \; SNR_i \tag{2.2}$$

where the SNR is decomposed into b one-third octave band contributions SNR_i and I_i is the weight for the importance of the i-th one-third octave band for speech intelligibility (Pavlovic, 1987). The performance of a noise reduction algorithm is assessed by measuring the gain in intelligibility weighted SNR

between the input, the signal of a microphone in our case, and the output of a noise reduction system:

$$G_{AI} = SNR_{AI,output} - SNR_{AI,input} \tag{2.3}$$

The transfer function between the input signal (In) (e.g. a microphone signal) and the output (Out) of the noise reduction technique can be estimated for the speech signal ($H_{s,\hat{s}}(f)$) and the noise signal ($H_{n,\hat{n}}(f)$):

$$H_{s,\hat{s}}(f) = \frac{P^{speech}_{Out,Out}(f)}{P^{speech}_{In,In}(f)} \quad H_{n,\hat{n}}(f) = \frac{P^{noise}_{Out,Out}(f)}{P^{noise}_{In,In}(f)} \tag{2.4}$$

where $P^{speech}(f)$ and $P^{noise}(f)$ are the PSD of the microphone signal (In) and of the output signal of the noise reduction strategy (Out) for the speech and the noise signals, respectively.

2.2.1.2 Physical evaluation

The performance measures for the physical evaluations are based on the computation of the SNR_{AI}, of polar diagrams and the directivity index (DI). The transfer function of the speech signal and the noise signal between the input and the output of the noise reduction strategy is also measured.

Since 1997, an extension of the articulation index calculation has been suggested. This extension is known as the Speech Intelligibility Index (SII) (ANSI, 1997), where the weights for the audibility function A_i are comprised of a variety of adverse listening conditions, such as noise masking, filtering, distortion and low reverberation. Based on the SII, the intelligibility weighted SNR becomes:

$$SNR_{SII} = \sum_{i=1}^{b} I_i \; A_i \; SNR_i \tag{2.5}$$

where A_i is calculated in accordance with the SII procedure at the i-th one-third octave band at the standard speech spectrum level for a raised vocal effort (68.3dBSPL) (ANSI, 1997). This speech spectrum level was chosen because of the sound power level of the loudspeaker during the recordings (70dBSPL). The gain G_{SII} in intelligibility weighted SNR between the input and the output of the noise reduction strategy is defined as:

$$G_{SII} = SNR_{SII,output} - SNR_{SII,input} \tag{2.6}$$

The only difference between the G_{SII}, used in the physical evaluation, and the G_{AI}, used in the theoretical evaluation, is the audibility function A_i. The

computation of the latter is based on the characteristics of the speech and the noise signals. However in the theoretical evaluation, the inputs of the noise reduction strategies are defined by the transfer function between the microphones (see section 3.2 and 4.2) and not by the signals of the speech and the noise themselves. Thus, the G_{SII} cannot be applied to the theoretical evaluation.

If the SNR_{SII} is calculated for different locations of a sound source, a polar diagram can be constructed. A polar diagram shows the sensitivity of the signals (e.g. omnidirectional microphone or the output of the noise reduction algorithms) as a function of the direction of the source (i.e. azimuth). The plotted values are relative to the direction 0^o (in front of the head), which is considered as the angle of incidence of the speech source.

The directivity index DI (Beranek, 1954) is often-used to measure the performance of a directional microphone configuration and noise reduction schemes in hearing aids (Desloge et al., 1997; Ricketts, 2000a). It has previously been shown that the DI has a strong link with the improvement of speech intelligibility in noise (Ricketts, 2000a). With θ the azimuth coordinate and ϕ the elevation coordinate, the directivity index equals:

$$DI(f) = \frac{4\pi \ |P(f,0,0)|^2}{\int_0^{2\pi} \int_0^{\pi} |P(f,\theta,\phi)|^2 \ |sin\theta| \ d\theta \ d\phi} \tag{2.7}$$

where the $|P(f,\theta,\phi)|^2$ is the mean squared sound pressure, at frequency f, of the output signal of the hearing aid when a single sound source is located at the coordinate (θ,ϕ). If symmetry is assumed in the vertical plane and there is reasonable symmetry around the horizontal plane, the DI can be calculated from only the $|P(f,\theta,\phi=0)|^2$ values recorded at discrete angles of the horizontal plane by using the following formula (Beranek, 1954):

$$DI(f_i^c) = \frac{8\pi \ 57.3^o \ |P(f_i^c,0,0)|^2}{2\pi \sum_{i=1}^{180^o/\delta\theta} |P(f_i^c,\theta_i,\phi=0)|^2 \ |sin\theta_i| \ \delta\theta_i} \tag{2.8}$$

where f_i^c is the center frequency of the ith one-third octave band. An intelligibility weighted version of the DI can be defined as:

$$DI_{AI} = \sum_{i=1}^{b} I_i \ DI(f_i^c) \tag{2.9}$$

where I_i is the weight for the importance of the intelligibility at the i-th one-third octave band (Pavlovic, 1987) and b the total number of one-third octave bands.

To analyze the influence of the noise reduction algorithms on the speech and the noise signal separately, the transfer functions $H_{s,\hat{s}}(f)$ and $H_{n,\hat{n}}(f)$, between the input and the output signals are calculated (equation 2.4).

2.2.1.3 Perceptual evaluation

For the perceptual evaluations, speech recognition tests are done with normal hearing and/or hearing impaired subjects. An adaptive procedure, comparable with the one used by Plomp and Mimpen (1979), is used. The adaptive procedure adjusts the level of the speech material to extract the 50% speech recognition level, i.e. the SRT. The noise level is kept constant during the test and a first sentence or word of a speech list is repeated with increasing level until the subject correctly identifies it. Once this level is determined, every other sentence or word of the list is subsequently presented only once at a level lower or higher, depending on the former item being identified correctly or not. These tests yield results accurate to within about 1dB and allow efficient determination of SRTs for the different conditions. The SRT-improvement is defined as the difference between the SRT obtained at a microphone signal (omnidirectional or directional microphone) and the SRT obtained at the output of a noise reduction strategy.

2.2.2 Voice activity detector

Many of the noise reduction techniques use the intermittent properties of the speech signal (**figure 2.2a**). Indeed, the statistics of the noise-only signals can be estimated during the *noise-only periods*, and then subtracted from the speech-and-noise signal during the *speech-and-noise periods*. To discriminate between the speech-and-noise periods from the noise-only periods, a Voice Activity Detector (VAD) is generally used (**figure 2.2b**). The basic function of a VAD is to compare the features of the incoming signal with thresholds, which are usually extracted from the characteristics of the noise and speech signals. In the sequel, an overview of different VADs is given.

• *Zero-crossing rate method*: This method consists of computing the zero-crossing rate (the average number of sign changes in the signal) of the input signal and comparing it with a preset threshold value (Savoji, 1989; Junqua et al., 1991). The zero-crossing rate of the noise signal is assumed considerably higher than the rate for the speech signal. This technique has problems in the presence of speech or noise signals with high zero-crossing rates.

• *Periodicity estimator method*: The objective of this method is to find the periodic component of the signal. Speech-and-noise periods are detected whenever the periodicities exceed a fixed threshold (Friedmann, 1978). However, the major problem in designing a VAD based on periodicity is its sensitivity to any periodic signal, which may also be an interference or background

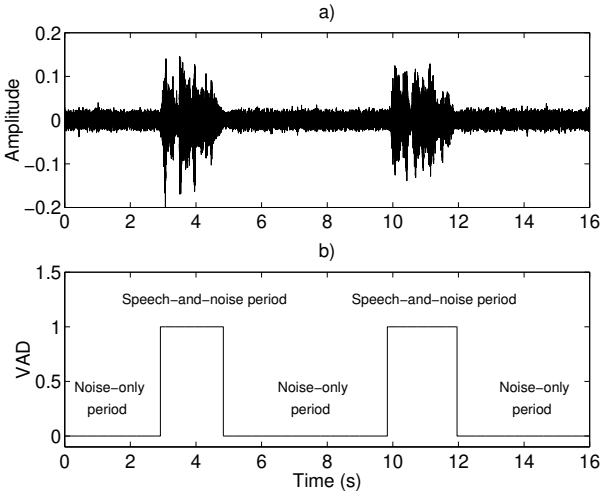


Figure 2.2: a) A noisy speech signal where b) a VAD detects the speech-and-noise periods and the noise-only periods.

noise signal. Tucker (1992) used a preprocessor to remove most of the interference. More recently, a VAD was proposed to detect the noise-only periods by adaptively tracking the minima of the temporal power envelope as well as the low and high-pass frequency band power envelopes of the signal (Marzinzik and Kollmeier, 2002).

• *Log-likelihood based method*: The decision rule of the VAD is derived from a likelihood ratio (LR) and the unknown parameters are estimated by using the maximum likelihood criterion (Sohn and Sung, 1998). However, important fluctuations of the estimated LR occur during noise-only periods. Improvements of this method were proposed where a smoother estimation of different parameters was performed by employing a decision-directed method (Sohn, 1999) or a first-order recursive filter (Cho et al., 2001).

• *Energy or magnitude based method*: The average energy or magnitude of the incoming signal is computed and the decision of the VAD is performed by comparing the energy estimate with threshold values. Generally, a first-order recursive filter is used to obtain a smooth estimate of the energy of the signal. The speech signal is assumed to be sufficiently larger than the background noise. Different variants of this method were proposed and depend on the computation and the choice of the thresholds. Vanden Berghe and Happaerts (1995) used the statistics (mean and variance) of the log-energy of the input signal to calculate the thresholds and to make the decision. This method was an extension of a principle developed by Van Gerven and Xie (1997). The methods developed

by Heitkamper (1997) and Savoij (1989) detected the speech-and-noise periods when the energy estimate of the signal was larger than the energy estimate of the noise signal. However, the estimate of the noise level was performed when a noise-only signal was present. Recently, a geometric adaptive energy threshold method was developed for the estimation of the noise level without the need for the speech pauses in the signal (Ozer and Tanyer, 1998; Tanyer and Ozer, 2000). This algorithm is based on the amplitude probability distribution of the signal. A modified amplitude probability distribution function (MAPD) of the input signal is computed by using a sorting operation. The speech and noise samples are partially separated on the MAPD plot and a geometric algorithm is proposed to estimate the noise level. The decision of the VAD is made by comparing the estimated noise level with the energy of the incoming signal.

• *Inter-microphone correlation based method*: The use of more than one microphone creates the opportunity to exploit the correlation between the microphones. The speech and the noise source are differentiated according to the angle of incidence of the sound (Greenberg and Zurek, 1992; Kompis et al., 1997; Beritelli et al., 1999). A speech signal detection scheme determines an opening angle, in which sources are treated as speech sources. The opening angle is approximately $\pm 20^o$ to allow head movement during conversation. However, these methods are difficult to apply for dual microphone hearing aids. Due to the small spacing between the microphones, the angle of incidence of the source cannot be defined accurately.

Tanyer and Ozer (2000) studied the performance of the most common VADs based on signal energy, zero-crossing rate and periodicity. The VADs discriminate well under high SNR and stationary noise conditions. In the case of low SNR and non-stationary noise, the performance drops significantly. Some studies evaluate the influence of VAD errors on the noise reduction systems (Le Bouquin and Faucon, 1995; Maj et al., 2002b; Doclo, 2003; Spriet et al., 2003a) where misclassification of speech-and-noise periods and noise-only periods may severely degrade the performance of the noise reduction algorithms, e.g. the SNR improvement and the signal distortion. Doclo (2003) showed that the VADs, based on the log-energy of the signals, affect the SNR improvement and the signal distortion less than the most common VADs for different types of noise.

A definition of VAD errors is given by Rosca et al. (2002) and to improve their performance, different strategies were proposed. Some of them are based on the combination of different methods (zero crossing, energy, periodicity) (Junqua et al., 1994; Kim et al., 2000; Tanyer and Ozer, 2000) or use a preprocessing stage to improve the SNR of the input signal (Lee and Ahn, 2001; Maj et al., 2002c; Rosca et al., 2002).

2.2.3 State-of-the-art noise reduction algorithms

The state-of-the-art noise reduction strategies are reviewed in this section. Different techniques are described and their contributions to the improvement of speech intelligibility in noise is examined. This section contains two parts. The first subsection describes the noise reduction algorithms implemented in commercial devices and the second describes the algorithms currently being investigated. In appendix A, a brief review of discrete-time signal processing is presented and only the definitions which are important for the signal processing considered in this thesis are mentioned.

2.2.3.1 Algorithms in commercial hearing aids

The development of noise reduction strategies in commercial devices is guided by the technical limitations of the hearing aids such as the housing of the devices, the calculation power of the DSPs and the number of microphones available.

2.2.3.1.1 High-pass filter

The first and the simplest noise reduction algorithm implemented in hearing aids was a high-pass filter. It was assumed that the energy of environmental noises was mainly at low frequencies. The hearing aid used a fixed high-pass filter which could be manually switched on or off as required. Afterwards, variable high-pass filters were used. These filters were adjusted to increasingly reject the energy of the low frequencies when this energy increased. In a non-noisy environment, the high-pass filter became an all-pass filter.

2.2.3.1.2 Multiple band-pass filters

Plomp (1988) argued that the speech intelligibility for hearing-impaired persons could be improved by a selective expansion or compression of the temporal modulation envelope of the signals under certain conditions. Based on this argument, Clarkson and Bahgat (1991) proposed a method which uses a filter bank to separate the input signal of the microphone into different frequency channels. The modulation frequency in each channel is analyzed to decide whether the signals are more likely to be speech or noise and each frequency range is amplified or attenuated accordingly. However, the noise reduction system did not produce any SRT-improvements (Alcantara et al., 2003).

2.2.3.1.3 Hardware directional microphone

High-pass and multiple band-pass filters work in the frequency domain. On the other hand, directional microphones work on the spatial characteristics of the sounds. They are designed to be most sensitive to sounds arriving from

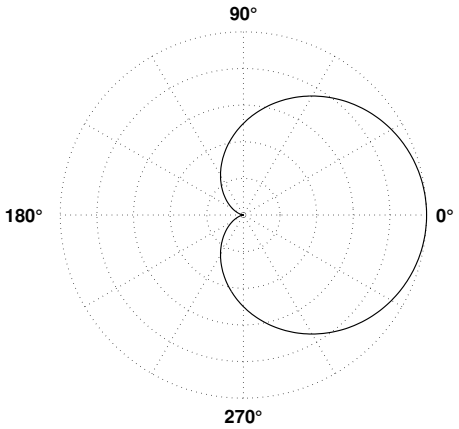


Figure 2.3: Cardioid spatial characteristics (null at 180^o) of the hardware directional microphone ($\beta = 1$).

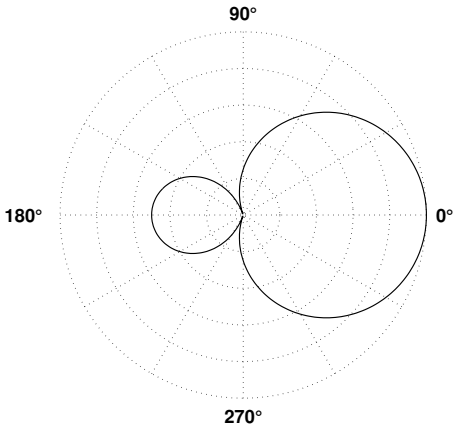


Figure 2.4: Hypercardioid spatial characteristics (null at 110^o) of the hardware directional microphone ($\beta = 3$).

the front and try to cancel (or to null) noise sounds coming from a specific direction. The directional microphone has two entry ports that allow sound to enter both the front and the rear cavities and arrive on either side of the microphone diaphragm. If the delayed version (internal delay) of the sound at the rear port reaches the diagram at the same time as the sound coming from the front port, a cancellation of the sound occurs. The spatial characteristic of the directional microphone depends on this formula:

$$D(\theta) = 1 + \beta \; cos(\theta) \tag{2.10}$$

where β is the ratio between the external delay (the travel time of sound due to distance between the two ports of the microphone) and the internal delay. The spatial characteristics of the directional microphone for different values of β are illustrated in **figure 2.3** and **2.4**. Several studies show that the directional microphone gives a SRT-improvement of about 3dB in difficult listening conditions (Hawkins and Yacullo, 1984; Leeuw and Dreschler, 1991; Maj et al., 2003).

2.2.3.1.4 Software directional microphone

Based on the hardware directional microphone strategy, software directional microphones have been developed (Bachler and Vonlanthen, 1995; Thompson, 1999). This method uses two omnidirectional microphones as the front and the rear entry port of the directional microphone (**figure 2.5**). The directional microphone signal is computed as the difference between the signal from the front microphone and the delayed-weighted signal of the rear

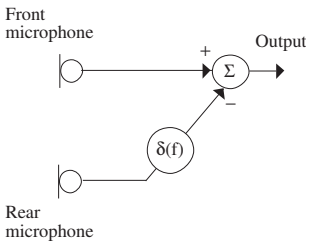


Figure 2.5: Representation of the software directional microphone.

microphone, resulting in a response comparable to a hardware directional microphone. The microphone parameter is the frequency-dependent weight for the back port $\delta(f) = a\ e^{-j2\pi f\tau}$ and the spatial characteristic of the directional microphone varies as a function of the internal delay τ and the weight a. Compared to the hardware directional microphone, similar improvements of the SRT (3-3.5dB) were found with the software directional microphone in reverberant conditions ($T_{60} \approx 0.45 - 0.76s$) (Wouters et al., 1999; Maj et al., 2003).

2.2.3.1.5 Adaptive directional microphone

The adaptive directional microphone, currently the state-of-the-art solution in modern commercial digital hearing aids, such as Phonak Claro and GNReSound Canta7, is illustrated in **figure 2.6** (Luo et al., 2002). Two software directional microphones provide reference signals, namely the *speech reference* and the *noise reference*. The speech reference is created with a front spatial cardioid (e.g. null at 180^o) and the noise reference is created with a rear spatial cardioid (e.g. null at 0^o). The signals of the software directional microphones, speech and noise reference, are connected to an Adaptive Noise Canceller (ANC) (Widrow and Stearns, 1985). The coefficients of this ANC filter can be updated by the means of classical adaptive algorithms (Haykin, 1996). Also, a constraint is applied on the coefficients of the ANC. This constraint allows the adaptation of the coefficient when a source is at the back hemisphere (where the source is considered to be noise) and stop the adaptation when a source is at the front hemisphere (where the source is considered to be speech). This avoids the cancellation of the speech signal at the output of the ANC.

Ricketts and Henry (2002) evaluated a software directional microphone and an adaptive directional microphone for hearing aids, in a moderately reverberant condition ($T_{60} \approx 400 - 450ms$). In this study, it was shown that the advantage of the adaptive over the fixed directional microphone was prominent when noise sources are situated on the side of the listener. With two noise

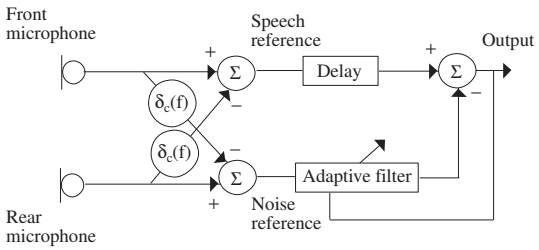


Figure 2.6: Representation of the adaptive directional microphone.

sources at 75^o and 105^o, an SRT-improvement of about 2dB was obtained with the adaptive directional microphone relative to an omnidirectional microphone. In this noise scenario, the software directional microphone gave roughly the same SRT as the omnidirectional microphone. With noise sources at 165^o and 195^o, both directional microphones had roughly the same behaviour and a SRT-improvement of about 4dB was obtained relative to the omnidirectional microphone.

2.2.3.2 Algorithms in research

In research, most of the solutions can afford to use a larger computing power (e.g. a PC) and the maximum intermicrophone distances are sometimes stretched beyond the dimensions of a single BTE hearing aid.

2.2.3.2.1 Power spectral subtraction and Wiener filter

The power spectral subtraction is a well-known single microphone approach. Several extensions were developed (Thomas and Niederjohn, 1968; Weiss, 1974; Boll, 1979; Lim and Oppenheim, 1979) and a generalization was introduced (Lim and Oppenheim, 1979):

$$|\hat{P}_{s,s}(f)|^h = |P_{u,u}(f)|^h - g\,|\hat{P}_{n,n}(f)|^h \tag{2.11}$$

The power spectral density (PSD) of the noise signal $\hat{P}_{n,n}(f)$ is approximated during noise-only periods and is then subtracted from the PSD of the speech-and-noise signal $P_{u,u}(f)$ during speech-and-noise periods to estimate the PSD of the clean speech $\hat{P}_{s,s}(f)$. It is assumed that speech and noise are uncorrelated and short-time stationary processes. To discriminate between speech-and-noise periods and noise-only periods of the input microphone signal, a VAD is generally used. When the constants g and h are set to unity, the equation (2.11) describes the system developed by Boll (1979). More complex variants were derived. McAulay and Malpass (1980) estimated the maximum likelihood of

$|P_{s,s}(f)|$ assuming that at each frequency the noise is Gaussian. Ephraim and Malah (1984) developed an algorithm based on a Minimum Mean-Square Error (MMSE). Spectral subtraction algorithms using the masking properties of the human auditory system were also developed (Virag, 1999; Li et al., 2001).

Another single microphone approach is based on the Wiener filtering and takes the form:

$$\hat{P}_{s,s}(f) = W_{WF}(f)\ P_{u,u}(f) \tag{2.12}$$

where $W_{WF}(f)$ is the frequency response of the generalized Wiener filter given by:

$$W_{WF}(f) = \left(\frac{\hat{P}_{s,s}(f)}{\hat{P}_{s,s}(f) + m\ \hat{P}_{n,n}(f)} \right)^n \tag{2.13}$$

So that,

$$|\hat{P}_{s,s}(f)| = \left(\frac{|\hat{P}_{s,s}(f)|}{|\hat{P}_{s,s}(f)| + m\ |\hat{P}_{n,n}(f)|} \right)^n |P_{u,u}(f)| \tag{2.14}$$

For some values of g, h in (2.11) and m, n in (2.14), the spectral subtraction and the Wiener filter algorithms are mathematically identical. For instance, if $n = 1/2$ and $h = 2$, the equation (2.14) is the same as (2.11). Also, if $n = 1$ and $m = 1/4$, the formula (2.14) is identical to the maximum likelihood estimate of McAulay and Malpass (1980).

With these single microphone techniques, an improvement in physical SNR was found, but unfortunately an improvement in SRT was not observed (Levitt et al., 1986). Furthermore, these strategies often introduce distortions on the signals, referred as 'musical noise'.

2.2.3.2.2 Beamformer

The beamformer technique is illustrated in **figure 2.7**. This strategy is designed to receive a signal radiating from a specific direction and attenuate signals radiating from other directions of no interest. In other words, the beamformer is constrained to pass the target signal with unit response:

$$\mathbf{w}(f)^T \mathbf{d}(f) = 1 \tag{2.15}$$

while at the same time minimizing the output power:

$$\underset{\mathbf{w}(f)}{min} \quad \mathbf{w}(f)^T \mathbf{R}_{u,u}(f)^{-1} \mathbf{w}(f) \tag{2.16}$$

where $\mathbf{w}(f)$ is the filter of the beamformer, $\mathbf{d}(f)$ is the propagation vector of the desired speech signal to be protected and $\mathbf{R}_{u,u}(f)$ is the $M \times M$ inter-microphone correlation matrix at the frequency f (M: number of microphone inputs). The *steering vector* $\mathbf{d}(f)$ corresponds to the transfer function between

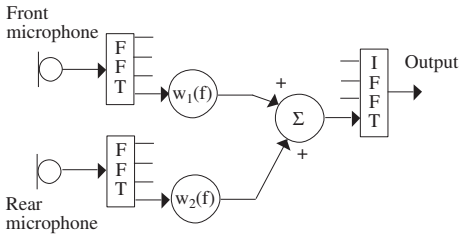


Figure 2.7: Block diagram of the beamformer.

each omnidirectional microphone in response to the plane waves arriving from the desired target, e.g. $\mathbf{d}(f) = [1 \quad e^{j2\pi fd/c}]^T$. The formula (2.15) and (2.16) constitute a standard linearly constrained optimization problem and a conventional Lagrange multipliers procedure leads to the solution:

$$\mathbf{w}(f) = \frac{\mathbf{R}_{u,u}(f)^{-1}\mathbf{d}(f)}{\mathbf{d}(f)^H\mathbf{R}_{u,u}(f)^{-1}\mathbf{d}(f)} \tag{2.17}$$

Different assumptions can be made for the correlation matrix $\mathbf{R}_{u,u}(f)$ and result in different beamformer strategies.

With **fixed beamformers**, the processing is time invariant. For the so-called delay-and-sum beamformer, the dominant noise is assumed to be the self-noise of the microphones (uncorrelated white noise) and not the ambient noise field. Since this Gaussian noise has equal intensity and is independent at each microphone, the correlation matrix $\mathbf{R}_{u,u}(f) = \mathbf{I}$ ($\mathbf{I}$ is the identity matrix) and hence the coefficients $\mathbf{w}(f) = \mathbf{d}(f)/M$. Soede et al. (1993a, 1993b) used the delay-and-sum strategy with an *endfire*[2] microphone array where five cardioid microphones were spaced on a 10cm-long array. A SRT-improvement of about 7dB was obtained in monaural conditions as compared with an omnidirectional microphone system. The perceptual evaluation was carried out in an anechoic room, where diffuse noise was simulated. Based on the study of Soede, Etymotic developed a prototype to be used as an assistive multimicrophone array for hearing aids. A physical and a perceptual evaluation was performed (Luts et al., 2003). The overall length of the array equaled 11.6cm and five bipolar microphones were used. Perceptual tests in noise were carried out in a reverberant room ($T_{60} \approx 0.76s$), with two types of noise (stationary speech weighted noise and multitalker babble) and six different noise source scenarios with single and multiple noise sources. Ten normal-hearing subjects and ten hearing aid users participated. Improvements up to 12dB for normal hearing and 9dB for hearing impaired listeners were obtained with three active array microphones relative to an omnidirectional microphone for one noise source at

[2] An endfire array has its microphones colinear with the target direction.

90^o. For three uncorrelated noise sources at 90^o, 180^o and 270^o improvements of about 9dB and 6dB were obtained for normal-hearing and hearing-impaired listeners, respectively. Even with a single noise source at 45^o benefits of 4dB were achieved. Five active microphones in the array could provide an additional improvement of about 1dB relative to the three-microphone configuration for a noise source at 45^o. Another type of fixed beamformer assumes that the ambient noise is isotropic (diffuse noise) and the intermicrophone correlation matrix $\mathbf{R}_{u,u}(f)$ for omnidirectional elements is a function of the type $sinc(2.\pi.f.d/c)$, where f is the sound frequency, c the velocity of the sound in the air ($c \approx 340m/s$) and d the intermicrophone spacing. Based on this assumption, Desloge et al. (1997) developed a microphone array for *binaural*[3] hearing aids. In a reverberant room ($T_{60} \approx 0.6s$), a benefit of 2-4dB was measured in a diffuse field, compared to a single cardioid microphone. Stadler and Rabinowitz (1993) studied the potential of different fixed beamformers in *broadside*[4] or in endfire orientation in terms of directivity (DI) for different types of microphones (omnidirectional, cardioid, dipole and hypercardioid). With a broadside orientation arrays based on cardioid and hypercardioid microphones gave similar performance (7-8dB). With an endfire orientation, similar results were obtained. However, with weights chosen independently for each frequency band, directivities up to 11dB might be obtained at the expense of increasing the noise sensitivities.

With **adaptive beamformers**, the coefficients $\mathbf{w}(f)$ and the intermicrophone correlation matrix $\mathbf{R}_{u,u}(f)$ are updated during processing (Frost, 1972). Errors in the microphone array characteristics, such as amplitude, phase responses and position can occur. If these errors are not correlated from sensor to sensor, they pass through the beamformer like uncorrelated or spatially white noise and can reduce the performance of the strategy. Hence, the gain of the array against white noise ('white noise gain') is a measure of robustness. Using the adaptive algorithm of Frost, Cox et al. (1987) developed an adaptive beamformer, called scaled projection algorithm, which maximizes the array gain G subject to a constraint on the white noise gain $G_w(f)$ ($\mathbf{R}_{u,u}(f) = \mathbf{I}$):

$$\underset{\mathbf{w}(f)}{max}\ G(f) = \frac{|\mathbf{w}(f)^H \mathbf{d}(f)|^2}{\mathbf{w}(f)^H \mathbf{R}_{u,u}(f) \mathbf{w}(f)} \tag{2.18}$$

subject to

$$G_w(f) = \frac{|\mathbf{w}(f)^H \mathbf{d}(f)|^2}{\mathbf{w}(f)^H \mathbf{w}(f)} = \gamma^2 \leqslant T_w \tag{2.19}$$

The constrained value γ^2 is chosen less than or equal to the maximum possible white noise gain T_w. A similar constraint was developed by Hoffman et al. (1994).

[3] A situation involving listening with two ears.
[4] A broadside array has its microphones perpendicular with the target direction.

A physical evaluation of three fixed and two adaptive beamformers was carried out (Kates and Weiss, 1996). This evaluation provided useful information for selecting an array-processing algorithm as function of performance. The two first fixed beamformers were the delay-and-sum beamformer and an oversteered array. The latter was similar to a delay-and-sum beamformer but the time delays used in combining the microphone output signals were greater than the acoustic propagation times between the microphones. The third fixed technique was the optimal superdirective beamformer (Cox et al., 1986). The two adaptive beamformers were the scaled projection algorithm (Cox et al., 1987) and an extension of this technique using a composite structured correlation matrix (Godara and Gray, 1989; Godara, 1991). For this evaluation, a microphone array of 10cm long with five omnidirectional microphones was used. The beamformer techniques were evaluated for adverse listening conditions in two different reverberant rooms ($T_{60} \approx 0.25s$ and $T_{60} \approx 0.6s$). The speech stimuli were presented at 0^o and the noise, which was multitalker babble, was presented at 60^o, 105^o, 180^o, 255^o and 300^o azimuth around the array. The delay-and-sum beamformer gave the worst performances. The oversteered strategy outperformed the latter technique by about 2dB. The optimal superdirective beamformer gave an additional gain of about 2dB compared to the oversteered scheme. The adaptive procedure with the composite structured correlation matrix gave approximately the same SNR improvement as the oversteered strategy and was outperformed by the scaled projection algorithm by about 1dB. Saunders and Kates (1997) evaluated perceptually the delay-and-sum beamformer and the superdirective processing in the same acoustic conditions and with the same microphone array as in the latter study. These two strategies were compared to a single omnidirectional microphone and a single cardioid microphone. It appeared that the single cardioid microphone was more effective than the delay-and-sum beamformer. As predicted in the study by Kates, the superdirective strategy always outperformed the other strategies.

2.2.3.2.3 Generalized Sidelobe Canceller

The Generalized Sidelobe Canceller (GSC) is depicted in **figure 2.8** and consists of a fixed beamformer, a blocking matrix and an ANC. The fixed beamformer (e.g. a delay-and-sum beamformer) is steered towards the desired signal and so creates the *speech reference*. The blocking matrix creates a *noise reference* where the presence of the desired signal is minimized. Generally, the Griffiths-Jim blocking matrix $\mathbf{B}$ is used (Griffiths and Jim, 1982) and with an endfire microphone array is set to:

$$\mathbf{B} = \begin{bmatrix} 1 & -e^{j2\pi fd/c} \end{bmatrix} \tag{2.20}$$

where d is the distance between the two microphones and c the velocity of the sound in air ($c \approx 340m/s$). In other words, with a two microphone approach the noise reference signal is the difference between the rear microphone signal and the delayed version of the front microphone signal. The speech reference and

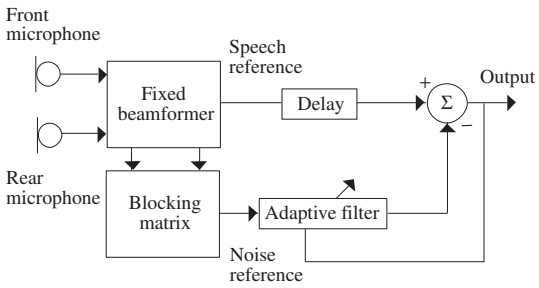


Figure 2.8: Block diagram for the Generalized Sidelobe Canceller.

the noise reference are the inputs of the ANC which minimizes the output power by moving the component in the speech reference signal that is correlated with the noise reference. The least mean square algorithm (LMS) is an often used procedure to update the coefficients of the ANC (Haykin, 1996). Buckley (1986) demonstrated that the GSC produces the same converged beamformer solution as the Frost beamformer (Frost, 1972). The GSC presents the advantages that the constraint in the adaptive process is conveniently implemented as a preprocessor and an intermediate conventional beamformer output is available at the speech reference signal.

A perceptual evaluation of the GSC was performed by ten normal hearing listeners (Peterson et al., 1987; Peterson, 1989). The desired signal was sent perpendicularly to the microphone array (broadside array) and the interference (multitalker babble) was separated by 45^o from the target. Speech intelligibility improvements of about 24dB in an anechoic room, were obtained over binaural listening (a microphone to each ear), and about 9dB and 0dB in two reverberant rooms, $T_{60} \approx 0.12s$ and $T_{60} \approx 0.48s$, respectively. Kompis and Dillier (2001a) proposed a theoretical method to predict the performance of the GSC. The prediction of the model was based on eleven different parameters including reverberation time and size of the room, directionality of the acoustic sources, and design parameters of the GSC itself (length of the adaptive filter and delay in the target signal path). An experimental verification of the theoretical predictions was provided (Kompis and Dillier, 2001b). The predictions agreed reasonably with the experimental results when the direct-to-reverberant ratio was smaller than 3dB, and underestimated the performance of the GSC for higher direct-to-reverberant ratios.

In the presence of errors in the microphone array characteristics such as microphone positions, microphone gains, reverberation and target direction, leakage of the desired signal in the noise reference can occur. Greenberg introduced a VAD and modified the LMS procedure of the ANC to avoid a can-

cellation of the desired signal (Greenberg and Zurek, 1992; Greenberg, 1998). Welker et al. (1997) assessed the strategy of Greenberg combined with the approach of Desloge et al. (1997) for a binaural output. The low frequencies (below 500Hz) of the spectrum were allocated for the binaural signals and the high frequencies (above 500Hz) were allocated for the GSC processing. The low frequency cues are important, especially the interaural delay, for binaural tasks. Measurements with normal hearing persons in a reverberant room ($T_{60} \approx 0.31s$) showed an improvement of approximately 3dB over simple binaural and 5dB over monaural listening conditions. A recent study (Greenberg et al., 2003) compared the approaches of Greenberg and Zurek (1992) and Desloge et al. (1997) with a delay-and-sum beamformer. On average the adaptive and the fixed beamformers provided SRT-improvements of about 7-9dB and 2dB respectively, over a binaural directional microphone in moderate reverberant conditions. Other studies designed the blocking matrix using adaptive filters to avoid the presence of the desired signal in the noise reference (Hoshuyama et al., 1999; Yu and Ueng, 2000). Van Compernolle (1990) modified the fixed beamformer of the GSC into an adaptive filter. This filter was only adapted when a VAD detected the presence of the desired signal. Vanden Berghe and Wouters (1998) evaluated the latter technique, which was called two-stage adaptive beamformer, for dual-microphone BTE hearing aids. In a moderately reverberant room ($T_{60} \approx 0.45s$), a 5dB perceptual improvement of speech understanding in noise was obtained, relative to the effect of one hardware directional microphone. The speech and the noise were presented at 0^o and 90^o azimuth (at the side of the hearing aid), respectively, relative to the position of the head.

2.2.3.2.4 Subspace and optimal filtering based noise reduction

The noise reduction algorithms based on subspace and optimal filtering reconstruct the time series of the speech signal $\mathbf{s}_k \in \mathbb{R}^{ML}$ (M is the number of the microphones and L is the length of the time series) at time k from noisy data $\mathbf{u}_k = \mathbf{s}_k + \mathbf{n}_k$ ($\mathbf{u}_k$ and $\mathbf{n}_k \in \mathbb{R}^{ML}$) by means of a linear filter $\mathbf{W}_k \in \mathbb{R}^{ML}$. Thus, the estimate of the speech signal $\hat{\mathbf{s}}_k$ takes the form:

$$\hat{\mathbf{s}}_k = \mathbf{W}_k^T \mathbf{u}_k \tag{2.21}$$

The $\hat{\mathbf{s}}_k$ is then an estimate for the speech contribution $\mathbf{s}_k$ in $\mathbf{u}_k$. The linear estimator $\mathbf{W}_k$ is obtained by using signal subspace or optimal filtering methods.

With **signal subspace** methods, the correlation matrix $\mathbf{R}_{u,u} \in \mathbb{R}^{ML \times ML}$ is decomposed[5] into a speech-and-noise part ($\mathbf{\Sigma}_s^j + \mathbf{\Sigma}_{n1}^j$) (the clean speech signal is modelled by a low-rank model of rank R) and a complementary noise part ($\mathbf{\Sigma}_{n2}^j$) (where j refers to the j-th microphone). In a two microphone approach,

[5] In appendix B, some definitions of linear algebra and matrix decompositions are given.

the correlation matrix $\mathbf{R}_{u,u}$ is decomposed as:

$$\mathbf{R}_{u,u} = \mathbf{R}_{s,s} + \mathbf{R}_{n,n} = \bar{\mathbf{X}} \begin{pmatrix} \mathbf{\Sigma}_s^1 + \mathbf{\Sigma}_{n1}^1 & 0 & \cdots & 0 \\ 0 & \mathbf{\Sigma}_{n2}^1 & \ddots & \vdots \\ \vdots & \ddots & \mathbf{\Sigma}_s^2 + \mathbf{\Sigma}_{n1}^2 & 0 \\ 0 & \cdots & 0 & \mathbf{\Sigma}_{n2}^2 \end{pmatrix} \bar{\mathbf{X}}^T \quad (2.22)$$

where $\mathbf{\Sigma}_s^1$, $\mathbf{\Sigma}_{n1}^1$, $\mathbf{\Sigma}_s^2$ and $\mathbf{\Sigma}_{n1}^2$ are $R \times R$-dimensional diagonal matrices, $\mathbf{\Sigma}_{n2}^1$ and $\mathbf{\Sigma}_{n2}^2$ are a $(L-R) \times (L-R)$-dimensional diagonal matrix. $\bar{\mathbf{X}} \in \mathbb{R}^{ML \times ML}$ is the matrix containing the eigenvectors, invertible but not necessarily orthogonal. The *singular value decomposition* (SVD) and the *eigenvalue decomposition* (EVD) are methods to decompose the (inter)microphone correlation matrix $\mathbf{R}_{u,u}$ (Haykin, 1996). Several criteria can be used to estimate the rank R of the speech-plus-noise part. Some of them are based on the statistics of the signal such as the Akaike Information Criterion (AIC) (Akaike, 1974) or the Minimum Description Length criterion (MDL) (Rissanen, 1978), and some of them are based on Gerschgorin disks (Caspary and Maj, 1999).

With **optimal filtering** methods, several linear estimators $\mathbf{W}_k$ are proposed in literature (Ephraim and Van Trees, 1995; Jensen et al., 1995; Hansen, 1997; Doclo and Moonen, 2002) and these can be formulated into an unified notation:

$$\mathbf{W}_k = \bar{\mathbf{X}}^{-T} \; diag\{f(\bar{\sigma}^2, \bar{\eta}^2)\} \; \bar{\mathbf{X}}^T \quad (2.23)$$

The formula (2.23) can be interpreted as follows: the matrix $\bar{\mathbf{X}}^{-T} \in \mathbb{R}^{ML \times ML}$ first performs a transformation from the time domain to a signal-dependent transform domain, and then, $f(\bar{\sigma}^2, \bar{\eta}^2)$ is a gain that is function of the generalized eigenvalues $\bar{\sigma}^2$ (speech-and-noise subspace) and $\bar{\eta}^2$ (noise subspace) and modifies the transform domain parameters. The matrix $\bar{\mathbf{X}}^T \in \mathbb{R}^{ML \times ML}$ finally performs the transformation back to the time domain.

The function $f(\bar{\sigma}^2, \bar{\eta}^2)$ depends on the specific cost criterion and the rank R of the speech model. The *Mean Square Error criterion* (MSE) minimizes the distance between the clean speech $\mathbf{s}_k$ and its linear estimate $\hat{\mathbf{s}}_k$. Thus, the linear estimator $\mathbf{W}_k$ is the well-known *Wiener filter* $\mathbf{W}_{WF}$:

$$\mathbf{W}_{WF} = \bar{\mathbf{X}}^{-T} \; \; diag\left\{ \frac{\bar{\sigma}_i^2 - \bar{\eta}_i^2}{\bar{\sigma}_i^2} \right\} \; \; \bar{\mathbf{X}}^T \quad (2.24)$$

The *Least Square criterion* (LS) minimizes the squared errors between the noisy speech signal $\mathbf{u}_k$ and a linear low-rank model of the clean speech signal $\mathbf{s}_k$. The linear estimator $\mathbf{W}_k$ becomes:

$$\mathbf{W}_{LS} = \bar{\mathbf{X}}^{-T} \; \mathbf{I}_R \; \bar{\mathbf{X}}^T \quad (2.25)$$

where $\mathbf{I}$ is the identity matrix of size R. The residual signal $\mathbf{e}_k = \mathbf{s}_k - \hat{\mathbf{s}}_k \in \mathbb{R}^{ML}$ minimized by the MSE or the LS methods represents the signal distortion $\mathbf{e}_s$ and the residual noise $\mathbf{e}_n$ as $\mathbf{e}_k = \mathbf{e}_{s,k} + \mathbf{e}_{n,k}$ (Hansen, 1997). Since both terms cannot be simultaneously minimized, Ephraim and Van Trees (1995) introduced a cost criterion which keeps the residual noise energy ϵ_n^2 under a threshold T_n while minimizing the energy of the signal distortion ϵ_s^2.

$$\min_{\mathbf{W}_k} \quad \epsilon_s^2 \quad subject \quad to \quad \epsilon_n^2 \leq T_n \quad where \quad 0 \leq T_n \leq 1 \tag{2.26}$$

Thus, the linear filter $\mathbf{W}_k$ becomes:

$$\mathbf{W}_E = \bar{\mathbf{X}}^{-T} diag\left\{ \frac{\bar{\sigma}_i^2 - \bar{\eta}_i^2}{\bar{\sigma}_i^2 + (\mu - 1)\bar{\eta}_i^2} \right\} \bar{\mathbf{X}}^T \tag{2.27}$$

The speech distortion parameter $\mu \in [0, \infty[$ (a Lagrange-multiplier) allows a trade-off between signal distortion and noise reduction. If $\mu = 1$ the original MMSE solution is obtained. More emphasis is put on the signal distortion when $\mu < 1$ at the expense of a decreased noise reduction performance. The residual noise level is reduced when $\mu > 1$ at the expense of increasing the speech distortion. With $\mu \to \infty$, all the emphasis is put on the noise reduction without taking into account the signal distortion. The cost criterion introduced by Ephraim was used with different decomposition methods for the (inter)microphone correlation matrix $\mathbf{R}_{u,u}$ such as the Karhunen-Loève Transform (KLT) transform (Ephraim and Van Trees, 1995), the Generalized-SVD (GSVD) (Doclo and Moonen, 2002) and the eigenvalue decomposition (Huang and Zhao, 1997). Mittal and Phamdo (2000) gave an extension of Ephraim's cost criterion and took into account the masking effect of the noise signal. Adaptive procedures were proposed to implement the method of Ephraim in real time (Rezayee and Gazor, 2001; Spriet et al., 2003c).

In a single microphone approach, only informal listening comparisons and subjective evaluations were carried out. In a multiple microphone approach, a noise reduction strategy developed by Doclo and Moonen (2002) based on the MSE criterion gave promising physical results and was seen to outperform the GSC strategy in adverse listening conditions (Spriet et al., 2001; Doclo, 2003; Spriet et al., 2004).

2.2.3.2.5 Blind source separation

Blind Source Separation (BSS) is a relatively new signal processing strategy, which was introduced by Herault and Jutten (1986). The goal of the BSS is to recover independent sources given only sensor observations that are linear mixtures of independent source signals. The term 'blind' indicates that both the source signals and the way the signals are mixed are unknown. There are two different methods to solve the BSS problem. A first is based on second-order statistic, called Principal Component Analysis (PCA). A second, called

Independent Component Analysis (ICA), does not only decorrelate the signals, such as PCA, but also reduces higher-order statistical dependencies. These techniques, PCA and ICA, have mainly been applied with success to biomedical signals, such as electroencephalography (EEG), electrooculography (EOG) and event-related potentials (ERP).

In speech processing, only PCA approaches were developed or evaluated. Two different types of signal separation can be distinguished, the scalar separation and the convolutive separation. On the one hand, the scalar separation assumes that the mixture between the sources is performed by a static mixing matrix and typical application areas are smallband signal processing. On the other hand, the convolutive separation assumes that the mixture between the signals is done by a transfer function and the sources are broadband like in speech and audio processing.

Nguyen-Thi (1992) evaluated both types of algorithms, scalar and convolutive separation, based on a PCA approach. The evaluation was carried out in anechoic condition and the sources, speech and noise, were close to the microphones. The SNR measure was used to evaluate the algorithms and the strategies based on the convolutive separation gave the best separation. Stability and non-unique solutions can occur with the separation algorithms. Van Gerven and Van Compernolle (1995, 1996) proposed an algorithm which was developed to tackle these problems. A convolutive separation strategy was used and applied to speech processing for two microphone applications. Schobben (1999) and Douglas and Sun (2003) also developed a convolutive separation algorithm for multiple microphone approaches. They used the time domain to update the coefficient of the filters, while Schobben used the frequency domain to get a faster convergence. A recent study also introduced a BSS algorithm using the frequency domain (Anemuller and Kollmeier, 2003). This algorithm was implemented in real time on a PC-platform anticipating its inclusion into the signal processing framework of digital hearing aids. The performance of the algorithm was evaluated both by simulations and experimentally, including the separation of a moving and a fixed speaker in an anechoic chamber. However, this strategy was limited to negligible acoustical reflections and needs improvement for a broader application in hearing aids. More extended overviews on the BSS can be found in (Lee, 1998) and (Douglas and Sun, 2003).

2.3 Conclusion

In this chapter, the properties of the signals recorded at the microphones of hearing aids, the signal processing basics and the performance measures considered in this thesis were described. An overview of VAD algorithms was given and state-of-the-art noise reduction strategies were examined.

In this thesis, two noise reduction strategies are optimized for dual microphone BTE hearing aids. The first algorithm, called two-stage adaptive beamformer (Vanden Berghe and Wouters, 1998), is an extension of the GSC procedure (section 2.2.3.2.3) while the second is an optimal filtering method for multimicrophone approaches (section 2.2.3.2.4) (Doclo and Moonen, 2002). Chapter 3 and 4 present the study of the two-stage adaptive beamformer and the optimal filtering strategy, respectively. Both noise reduction strategies are assessed theoretically, physically and perceptually.

The *theoretical evaluation* consists of the derivation of the equations of the noise reduction strategies. These equations, based on the complex coherence function between two microphones, describe the transfer function between the input and the output of the algorithms. Thus, a preliminary evaluation of the noise reduction technique can be performed by varying parameters such as the distance between the two microphones, the internal noise of the hearing aid, the characteristics of the microphones (gain and phase) and the VAD errors. The influence of these parameters on the noise reduction algorithms can be assessed by using the transfer functions $H_{s,\hat{s}}$ and $H_{n,\hat{n}}$ and the gain in intelligibility weighted SNR, G_{AI}. The G_{AI} is used in the theoretical evaluation because the G_{SII}, used in the physical evaluation, cannot be applied properly (section 2.2.1.2).

For the *physical evaluations*, the signals of the microphones from BTE hearing aids are recorded in different acoustic environments. The different parameters (e.g. length of the filter operations) of the noise reduction techniques are optimized by maximizing the speech intelligibility in noise, through the G_{SII}. The performance of the optimized strategies is evaluated by measuring the directivity index, the polar diagrams and the transfer functions $H_{s,\hat{s}}$ and $H_{n,\hat{n}}$ for different jammer sound source scenarios and acoustic environments.

For the *perceptual evaluations*, the noise reduction strategies are implemented in real time on different platforms (research or commercial, depending on the study). The adaptive procedure is used to estimate the SRT in noisy conditions and the SRT-improvements obtained by the noise reduction strategies relative to a microphone signal are evaluated. Normal hearing listeners and hearing impaired subjects perform the SRT measures in different jammer sound source scenarios in moderately reverberant conditions.

Chapter 3

Two-stage adaptive beamformer

In this chapter, a two-stage adaptive beamformer technique is developed and assessed theoretically, physically and perceptually. The aim of this study is to optimize this technique for a real time implementation in the most recent digital hearing aids and to bring an additional gain in speech intelligibility relative to state-of-the-art noise reduction algorithms in modern devices.

The *first section* describes the two-stage adaptive beamformer strategy. The theoretical analysis is carried out in the *second section* and the influence of different parameters, such as the distance between the microphones and the internal noise of the sensors, on the noise reduction performance of the strategy is examined. In the *third section*, the optimization of the adaptive beamformer for a real time implementation on a wearable research platform is performed. This optimization is validated physically and perceptually. In the *fourth section*, the noise reduction strategy is evaluated more extremely, for different microphone configurations and with three different jammer sound scenarios. In total, 1368 SRT-measurements are performed with normal hearing listeners and hearing impaired subjects. Finally in the *fifth section*, the performance of the two-stage adaptive beamformer is compared to an adaptive directional microphone. Both algorithms are implemented in the same commercial hearing aids. A physical experiment is carried out to evaluate the influence of an individualized calibration procedure on fixed processing (e.g. software directional microphone). A perceptual evaluation is performed to compare the performance of the two-stage adaptive beamformer relative to an adaptive directional microphone when both strategies are implemented in the same digital hearing aid.

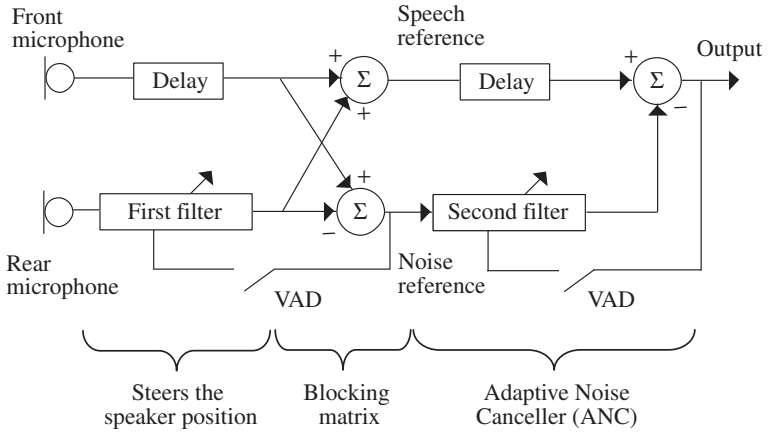


Figure 3.1: Scheme of the two-stage adaptive beamformer strategy introduced by Van Compernolle.

3.1 Two-stage adaptive beamformer

The two-stage adaptive beamformer (A2B) is an extension of the GSC procedure (section 2.2.3.2.3) and was introduced by Van Compernolle (1990) (**figure 3.1**). This strategy was applied to a BTE hearing aid and a significant gain in SRT of 5dB was found between a hardware directional microphone and the output of the two-stage adaptive beamformer (Vanden Berghe and Wouters, 1998). The processing consisted of two adaptive filters which were controlled by a VAD. The VAD was implemented to decide whether a speech-and-noise period or a noise-only period was present. The first filter adapted during speech-and-noise periods and was kept fixed during noise-only periods. This filter was used to follow the direction of the desired signal, e.g. under slowly moving target and head movements. The sum-and-subtraction stage was derived from the Griffiths-Jim blocking matrix and attempted to create the speech reference and the noise reference. The combination of the first adaptive filter and the blocking matrix limited the presence of speech in the noise reference and improved the noise signal estimate. The speech reference and the noise reference were then connected to an ANC, which attempted to model the noise contribution in the speech reference during noise-only periods, and subtracted noise from the speech reference during speech-and-noise periods.

The difference between Van Compernolle's approach and the GSC strategy resides mainly in the implementation of the first processing part. Van Compernolle's approach uses an adaptive filter which takes into account the errors of the microphone arrays and the position of the speaker, while the GSC

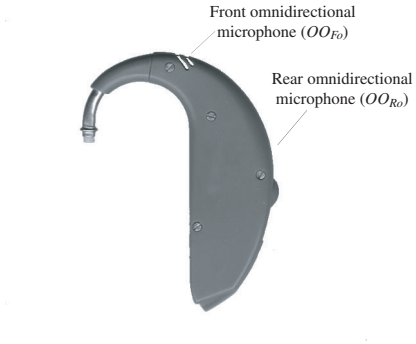


Figure 3.2: Danavox 163D hearing aid (OO) with two omnidirectional microphones mounted in an endfire array configuration spaced two centimeters apart.

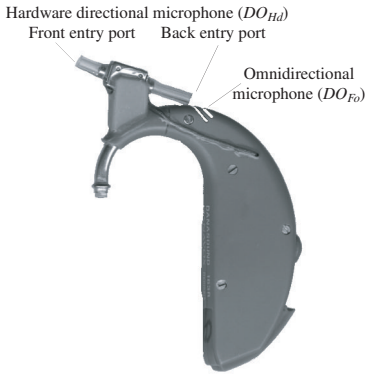


Figure 3.3: Danavox 163D hearing aid (DO) with one hardware directional microphone and one omnidirectional microphone mounted in an endfire array configuration.

procedure uses a fixed beamformer which assumes that no error arises in the microphone array and that the speaker is at a well-defined position. However, these procedures may have robustness problems. On the one hand, the technique with the first adaptive filter may converge to a wrong direction when different speakers are around the listener or if loud noise is wrongly detected as speech by the VAD. On the other hand, the fixed beamformer of the GSC procedure does not take into account the errors in the microphone array (e.g. microphone positions, microphone characteristics or reverberations) and this can lead to a partial cancellation of the desired speech.

In this study, the first filter of the two-stage adaptive beamformer is kept fixed and is optimized for each hearing aid. This solution is a trade-off between the adaptive filter of Van Compernolle's approach and the fixed beamformer of the GSC procedure. Indeed, the errors of the microphone arrays are taken into account to minimize the presence of the speech signal in the noise reference of the A2B. The filter is optimized for each hearing aid and the speaker is assumed to be in front of the listeners. A 0^o look direction is given to the two-stage adaptive beamformer.

To evaluate the performance of the A2B, two types of BTE hearing aids were used, referred to as OO (**figure 3.2**) and DO (**figure 3.3**). They had two different microphone configurations. Both hearing aids had a Danavox-163D housing, linear amplification and no compression or feedback control systems.

The OO type had two omnidirectional microphones (Knowles EM4368) which were mounted in an endfire array configuration spaced two centimeters

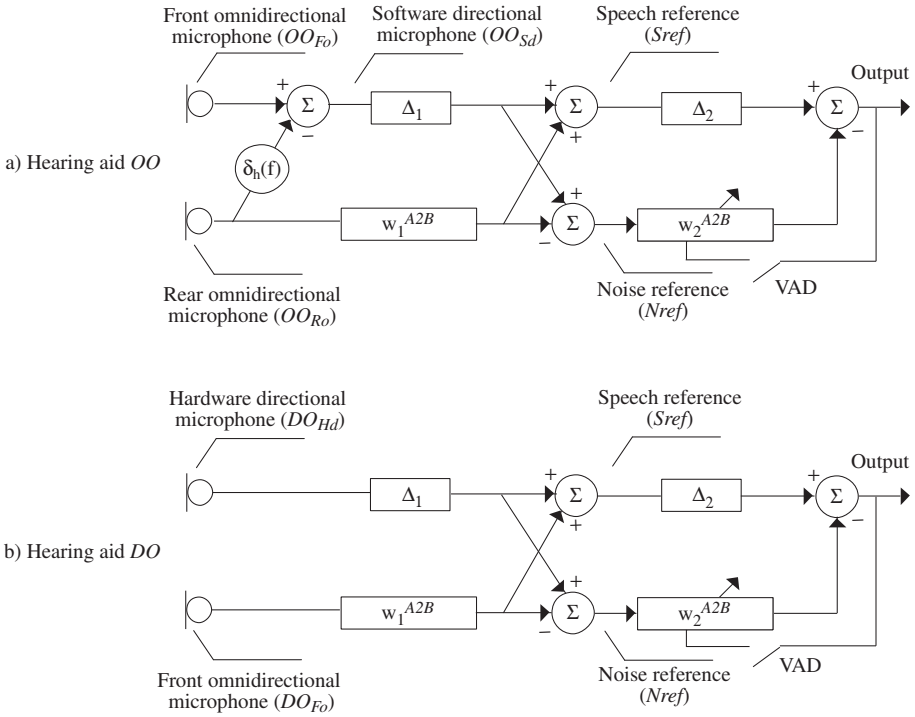


Figure 3.4: Representation of the two-stage adaptive beamformer (A2B) for the two types of hearing aids OO and DO. a) With hearing aid OO, two omnidirectional microphones (OO_{Fo} and OO_{Fo}), spaced by two centimeters, create a software directional microphone signal OO_{Sd}. The rear omnidirectional microphone and the software directional microphone are connected to the A2B. b) With hearing aid DO, one hardware directional microphone DO_{Hd} and one omnidirectional microphone DO_{Fo} are connected to the A2B.

apart (**figure 3.2**). The front (OO_{Fo}) and the rear (OO_{Ro}) omnidirectional microphones were used to create a software directional microphone (OO_{Sd}) (**figure 3.4**). The microphone parameter is the frequency-dependent weight for the back port $\delta(f) = a\ e^{-j2\pi f\tau}$ where the delay τ and the weight a were chosen to give a hypercardioid spatial characteristic in anechoic conditions (null at 110^o).

The DO type had one hardware directional microphone (DO_{Hd})(Knowles EL3077) and one omnidirectional microphone (DO_{Fo})(Knowles EL3346) (**figure 3.3**). The hardware directional microphone had a cardioid spatial characteris-

tic (null at 180°) in anechoic condition. The hardware directional microphone and the omnidirectional microphone signal were the inputs to the two-stage adaptive beamformer (**figure 3.4**).

The front omnidirectional microphone of OO_{Fo} and the omnidirectional microphone of DO_{Fo} had the same position on the hearing aid (**figures 3.2** and **3.3**). These omnidirectional microphones were the references for the different evaluations.

3.2 Theoretical Analysis

In this section, the theoretical evaluation of the two-stage adaptive beamformer is carried out. The equations of the transfer function between the input and the output of the noise reduction technique are derived for the simplified case where the number of taps in $\mathbf{w}_1^{A2B}$ and $\mathbf{w}_2^{A2B}$ is infinitely large. These equations are based on the complex coherence function and allow to assess the A2B theoretically. The scheme of the A2B applied to the hearing aid OO is used (**figure 3.4a**), and for brevity, the microphone indices OO are omitted in the equations. The scheme of the two-stage adaptive beamformer applied to the hearing aid DO is illustrated **figure 3.4b** and is used for the physical and the perceptual evaluations. The publications related to this section are (Maj et al., 2001a; Maj et al., 2002b).

The theoretical performance of the A2B was analysed by varying different parameters. These parameters were the distance between the microphones, the sensor-to-environmental noise ratio and the signal-to-noise ratio (SNR at the omnidirectional microphone). The G_{AI} was used in the theoretical evaluation because the G_{SII}, used in the physical evaluation, could not be applied properly (section 2.2.1.2).

3.2.1 Derivation of the equations

The CCF for the OO device is based on the (cross) power spectral densities (between) of omnidirectional microphone signals and is given by:

$$\Gamma_{Fo,Ro}(f) \triangleq \frac{P_{Fo,Ro}(f)}{\sqrt{P_{Fo,Fo}(f)\ P_{Ro,Ro}(f)}} \tag{3.1}$$

When a source is located at angle θ the CCF equals

$$\Gamma_{Fo,Ro}(f) = exp(j\ 2\pi\ f\ cos(\theta)\ d/c) \tag{3.2}$$

where d is the distance between the two microphones and c the velocity of the sound in air ($c \approx 340m/s$). The noise coherence is given by (Bitzer et al., 2001;

Spriet et al., 2004):

$$\Gamma^{noise}_{Fo,Ro}(f) = \frac{\Gamma^{env}_{Fo,Ro}(f)}{1+\rho(f)} \tag{3.3}$$

where $\Gamma^{env}_{Fo,Ro}(f)$ is the coherence function of the environmental noise (env) and $\rho(f) = P_{sens,sens}(f)/P_{env,env}(f)$ the sensor-to-environmental noise ratio. The environmental noise is the acoustic noise from the environment that is sensed by the microphone. The sensor noise ($sens$), or internal noise, in a microphone is the small amount of noise that is generated by the microphone or its internal preamplifier circuitry, and is unrelated to environmental noise. In this section, the speech source is located at the endfire angle 0^o and the noise source at the broadside angle 90^o. Thus, the CCFs for the speech and the noise sources are:

$$\Gamma^{speech}_{Fo,Ro}(f) = exp(j\ 2\pi\ f\ d/c) \quad \Gamma^{noise}_{Fo,Ro}(f) = \frac{1}{1+\rho(f)} \tag{3.4}$$

In the next paragraphs, the PSD at the output of the software directional microphone and the noise reduction algorithm are defined. It is assumed that the PSDs of the received signals at the microphone inputs are the same: $P^{noise}_{Ro,Ro}(f) = P^{noise}_{Fo,Fo}(f) \triangleq P^{noise}_{In,In}(f)$ and $P^{speech}_{Ro,Ro}(f) = P^{speech}_{Fo,Fo}(f) \triangleq P^{speech}_{In,In}(f)$. Finally, $\alpha(f)$ is defined as the signal-to-noise ratio at the omnidirectional microphones $P^{noise}_{In,In}(f) = \alpha(f)\ P^{speech}_{In,In}(f)$.

The PSD and the C-PSD as a function of the CCF at the output of the software directional microphone (OO_{Sd}) are:

$$\begin{aligned} P_{Sd,Sd}(f) &= P_{in,in}(f)\ (1 - 2\ Re(\delta^*_h(f)\ \Gamma_{Fo,Ro}(f)) + \delta_h(f)\ \delta^*_h(f)) \\ P_{Sd,Ro}(f) &= P_{in,in}(f)\ (\Gamma_{Fo,Ro}(f) - \delta_h(f)) \end{aligned} \tag{3.5}$$

The first filter $w^{A2B}_1(f)$ is kept fixed, under the assumption that the speaker is always in front of the listener at the angle 0^o. The filter equals:

$$w^{A2B}_1(f) = \frac{P^{speech}_{Sd,Ro}(f)}{P^{speech}_{Ro,Ro}(f)} \tag{3.6}$$

and the PSD and the C-PSD at the speech reference ($Sref$) and the noise reference ($Nref$) are:

$$\begin{aligned} P_{Sref,Sref}(f) &= P_{Sd,Sd}(f) + |w_1(f)|^2\ P_{Ro,Ro}(f) \\ &\quad + 2\ Re(w^*_1(f)\ P_{Sd,Ro}(f)) \\ P_{Nref,Nref}(f) &= P_{Sd,Sd}(f) + |w_1(f)|^2\ P_{Ro,Ro}(f) \\ &\quad - 2\ Re(w^*_1(f)\ P_{Sd,Ro}(f)) \\ P_{Sref,Nref}(f) &= P_{Sd,Sd}(f) - |w_1(f)|^2\ P_{Ro,Ro}(f) \\ &\quad + 2\ Im(w^*_1(f)\ P_{Sd,Ro}(f)) \end{aligned} \tag{3.7}$$

The second filter $w_2^{A2B}(f)$ is adapted during noise-only periods and equals:

$$w_2^{A2B}(f) = \frac{P^{noise}_{Sref,Nref}(f)}{P^{noise}_{Nref,Nref}(f)} \tag{3.8}$$

Finally, the PSD at the output of the two-stage adaptive beamformer is:

$$\begin{aligned} P_{Out,Out}(f) = P_{Sref,Sref(f)} + |w_2^{A2B}(f)|^2 \; P_{Nref,Nref}(f) \\ - 2 \; Re(w_2^{A2B*}(f) \; P_{Sref,Nref}(f)) \end{aligned} \tag{3.9}$$

3.2.2 Results and discussion

Simulations were carried out by varying the different parameters such as the distance d between the two microphones , the sensor-to-environmental noise ratio $\rho(f)$ and the signal-to-noise ratio $\alpha(f)$ when a speech source is present at 0^o and a noise source at 90^o.

The distance d between the two microphones had an influence on the transfer function of the speech signal $H_{s,\hat{s}}(f)$. The transfer functions $H_{s,\hat{s}}(f) = P^{speech}_{Out,Out}(f)/P^{speech}_{In,In}(f)$ and $H_{n,\hat{n}}(f) = P^{noise}_{Out,Out}(f)/P^{noise}_{In,In}(f)$ between the front omnidirectional microphone OO_{Fo} and the output of the two-stage adaptive beamformer as a function of the distance d between the two microphones for $\alpha(f) = 0$dB and $\rho(f) = -45$dB are illustrated in **figure 3.5**. Knee points could be defined where the magnitude of $H_{s,\hat{s}}(f)$ equaled 0dB. These knee points were at 2kHz, 3kHz and 4.2kHz for d=2cm, d=1.5cm and d=1cm, respectively. Below these knee points, $H_{s,\hat{s}}(f)$ was attenuated and the shorter the distance d the bigger the attenuation. Above these knee points, $H_{s,\hat{s}}(f)$ was amplified and the longer the distance d the bigger the amplification. With d=0.5cm, the speech signal was always attenuated ($H_{s,\hat{s}}(f) < 0$dB) from 0Hz to 8kHz. The transfer function between the software directional microphone and the output of the A2B is depicted in **figure 3.6**. Independently of the frequency, $H_{s,\hat{s}}(f)$ equaled 0dB for the different values of d. This means that the speech signal was not distorted by the processing of the two-stage adaptive beamformer but by the directional microphone. Thus, a small spacing allows an easier packaging of the microphones inside the hearing aids but affects the sensitivity of the software directional microphone at low frequencies. Thompson (1999) argued that a spacing of 12mm is a compromise for a software directional microphone. The longer the distance between the two microphones, the higher the G_{AI} (**figure 3.7**).

The internal noise in a microphone array depends mainly on the quality of the electronics and the microphones used in the hearing aids. **Figure 3.8** shows the influence of the sensor-to-environmental noise ratio $\rho(f)$ on the G_{AI} and the higher the sensor-environmental noise ratio $\rho(f)$, the worse the G_{AI}. Finally, the G_{AI} is independent of the signal-to-noise ratio $\alpha(f)$ (**figure 3.9**).

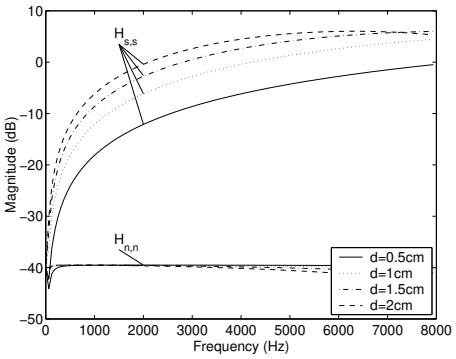


Figure 3.5: Influence of the distance between the two microphones on the transfer function of the speech signal $H_{s,\hat{s}}(f)$ and the noise signal $H_{n,\hat{n}}(f)$ between the front omnidirectional microphone and the output of the A2B ($\alpha(f) = 0$dB, $\rho(f) = -45$dB).

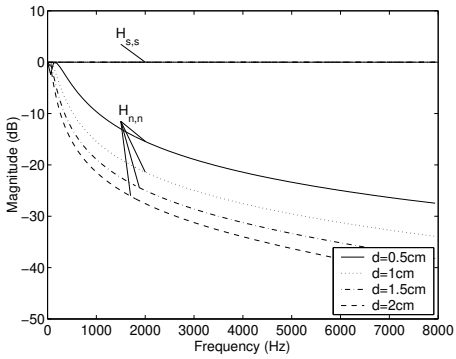


Figure 3.6: Influence of the distance between the two microphones on the transfer function of the speech signal $H_{s,\hat{s}}(f)$ and the noise signal $H_{n,\hat{n}}(f)$ between the software directional microphone and the output of the A2B ($\alpha(f) = 0$dB, $\rho(f) = -45$dB).

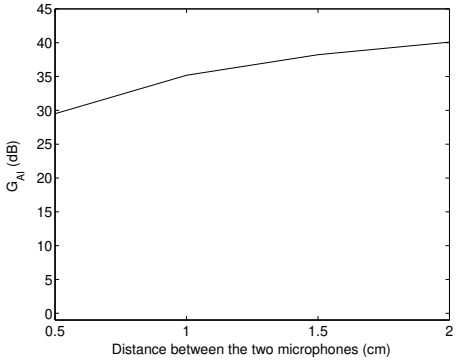


Figure 3.7: Influence of the distance between the two microphones d on G_{AI} ($\rho(f) = -45$dB, $\alpha(f) = 0$dB).

3.3 Optimization of the algorithm

This section presents the optimization of the two-stage adaptive beamformer for a real time implementation in a wearable research platform. This optimization was carried out with the hearing aid *DO* (**figure 3.3** and **3.4b**) and was validated physically and perceptually. The publication related to this section is (Wouters et al., 2002).

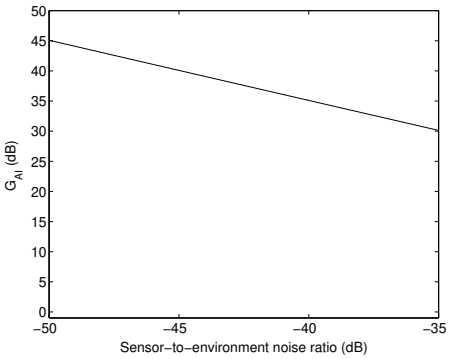


Figure 3.8: Influence of the sensor-to-environmental noise $\rho(f)$ on G_{AI} (d=2cm, $\alpha(f) = 0$dB).

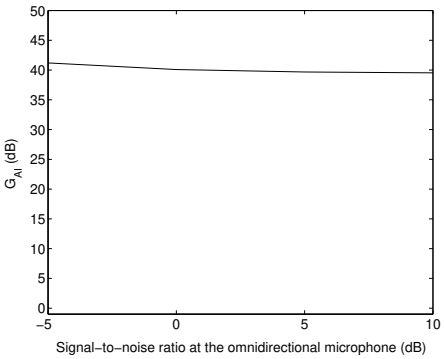


Figure 3.9: Influence of the signal-to-noise ratio $\alpha(f)$ on G_{AI} (d=2cm, $\rho(f) = -45$dB).

3.3.1 Implementation of the A2B

The implementation of the A2B consists of three parts: the determination of the coefficients of the first filter $\mathbf{w}_1^{A2B}$, the update process of the adaptive filter $\mathbf{w}_2^{A2B}$ in real time and the VAD algorithm.

3.3.1.1 First stage

To determine the coefficients of the fixed filter, an unmodulated signal based on the multilanguage long-term average speech spectrum (LTASS) from the 'ICRA' compilation was used (ICRA, 1997). This ICRA signal is white noise filtered in close accordance with the LTASS (Byrne et al., 1994). The ICRA signal used was an unmodulated noise representative of a male weighted idealized speech spectrum with a normal effort (ICRA, 1997). In anechoic conditions[1], this signal was presented by a loudspeaker one meter in front of a mannequin (at the angle 0^o) with a hearing aid (**figure 3.10**). The two microphones of the hearing aid were connected to equipment in a second room. The signals from the microphones were amplified with a two-channel Larson Davis 2200C amplifier and were digitized at a sampling frequency of 16kHz and a resolution of 16 bit on a PC-platform using a Texas Instruments TMS320-C40 DSP. The signals were downsampled from 16kHz to 15.625kHz (sampling frequency of the real time platform). The coefficients of the first filter $\mathbf{w}_1^{A2B}$ were determined by using the digitized signal under MATLAB and the Wiener-Hopf equation:

$$\mathbf{w}_1^{A2B} = \mathcal{E}\{\mathbf{do}_{Fo,k}\ \mathbf{do}_{Fo,k}^H\}^{-1}\mathcal{E}\{\mathbf{do}_{Fo,k}\ \Delta_1\mathbf{do}_{Hd,k}^H\} \tag{3.10}$$

[1] Anechoic chamber at the physics department of the KULeuven.

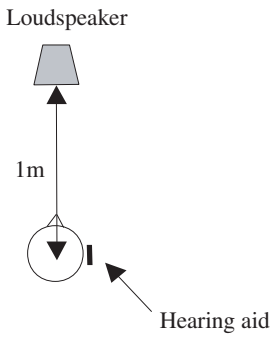


Figure 3.10: Set-up for the recordings of the microphone signals when the hearing aid is positioned on a dummy head.

where $\mathbf{do}_{Fo,k}$ is a time series of the omnidirectional microphone signal DO_{Fo} and $\Delta_1\mathbf{do}_{Hd,k}$ the delayed version of a time series of the directional microphone signals DO_{Hd} at time k (**figure 3.4**). The obtained coefficients were used to create the fixed first filter. The additional delay Δ_1 enabled the use of a non-causal filter, and its value was set to half of the size of the filter.

3.3.1.2 Second stage

The coefficients of the second filter were updated during real time processing by a normalized least mean squares procedure (NLMS) (Haykin, 1996):

$$\mathbf{w}_{2,k+1}^{A2B} = \mathbf{w}_{2,k}^{A2B} + \frac{\nu}{\varsigma + \|\mathbf{n_{ref}}_{,k}\|^2}\,\mathbf{n_{ref}}_{,k}\,(\Delta_2 s_{ref,k} - \mathbf{w}_{2,k}^{A2B\ H}\,\mathbf{n_{ref}}_{,k}) \quad (3.11)$$

where ν is the adaptation constant ($\nu = 0.06$), ς a positive constant, $\|\mathbf{n_{ref}}_{,k}\|^2$ the Euclidean norm of a time series of $\mathbf{n_{ref}}_{,k}$ (noise reference) at time k and $s_{ref,k} \in \mathbb{R}^1$ (speech reference) a sample of the desired response at time k. The additional delay Δ_2 also enabled the use of a non-causal filter, and its value was set to half of the size of the filter.

3.3.1.3 Voice Activity Detector

The VAD used in this study is based on the log-energy of the directional microphone signal[2] (hardware or software). Several extensions (Vanden Berghe and Happaerts, 1995; Doclo and De Clippel, 1997) of the VAD proposed by Van Gerven (1997) were developed and an evaluation of these different methods was carried out. The time constant and the sensitivity of the technique proposed by

[2] In this chapter, the VAD is always connected to the signal of the directional microphone (hardware or software depending on the type of hearing aid used).

Vanden Berghe and Happaerts (1995) to detect the speech-and-noise periods and the noise-only periods was satisfactory. The latter method is used in this thesis.

The log-energy of the signal was computed with an overlap method of 128 samples. The VAD brought only a delay of 2ms at 15.625kHz and had a computational cost of twenty three operations (additions or multiplications) per sample. The decision of the VAD was based on the computation of two thresholds, namely *Tspeech* and *Tnoise*, where *Tspeech* and *Tnoise* were computed from the statistics of the signal (the mean and the variance). The first function, *Signal*, equaled the log-energy when the energy of the signal increased, and dropped with an exponential curve when the energy dropped. The second function, *Offset*, preserved the *VAD=1* during a number of samples when a noise-only period was detected. In this way, a speech-and-noise period was still identified when there was a silence within a word or a sentence. With these different thresholds, the VAD worked as follows:

- if *Signal* > *Tspeech*, a speech-and-noise period was detected, *VAD=1*.
- if *Tnoise* > *Signal* and *Offset=1*, a noise-only period was detected but *VAD=1*.
- if *Tnoise* > *Signal* and *Offset=0*, a noise-only period was detected, *VAD=0*.

The thresholds of the VAD algorithm were defined for providing a good discrimination between the speech-and-noise periods and the noise-only periods with the software directional microphone signal at a SNR of 5dB in stationary and non-stationary conditions (Maj et al., 2003; Maj et al., 2003a; Maj et al., 2003b).

3.3.2 Physical evaluation

The physical evaluation was carried out in moderate reverberant conditions. The optimization of the size of the filters is performed by minimizing the SNR in the noise reference and maximizing the G_{SII} between the omnidirectional microphone and the output of the A2B. The polar diagrams of the front omnidirectional microphone signals, the directional microphone signals and the output signals of the A2B were measured. Furthermore, the transfer function between the directional microphone and the output of the strategy was estimated.

3.3.2.1 Experimental set-up

The test room used for the physical evaluations was a moderately reverberant room ($6 \times 3.75 \times 3.15 \mathrm{m}^3$, $T_{60} \approx 450ms$). The estimated direct-to-reverberant ratio was 7dB at 1m (Peutz, 1971). In this room, the hearing aid was mounted on the right ear of a dummy head. Speech and noise materials

were presented with identical loudspeakers (Yamaha CBX-S3) at one meter and at the same height as the center of the dummy head. The speech source was at 0^o azimuth and the azimuth angle of the noise source was varied (both sources at 70dBSPL at one meter). The two microphones of the hearing aid were connected to equipment in a second room and the signals were digitized with the same devices as described above (section 3.3.1.1). The test stimulus, fed to the loudspeakers, was a stationary speech-weighted noise derived from the sentences of Plomp and Mimpen (Plomp, 1979).

To carry out the physical evaluations, the coefficients of the second filter were fixed to the converged values when only noise signals were used. This means that the noise reduction algorithm was not in transient conditions and a perfect VAD was simulated. The same procedure was repeated with speech-weighted noise presented separately at different angles to calculate the respective attenuations. The levels of all signals presented were calibrated for a 0dB SNR at the center of the dummy head. By subsequent analysis of these recorded signals, the SNR of the omnidirectional microphone, directional microphone and at the output of the two-stage adaptive beamformer, were obtained as a function of angle of the jammer sound source. The different locations of the noise source corresponded to angles between 0^o and 345^o, in 15^o steps, relative to the direction of the speech source (0^o).

The aim of the first stage is to minimize the presence of the speech signal in the noise reference of the A2B and the aim of the second stage is to maximize the improvement of the speech intelligibility at the output of the A2B. Thus, the SNR and the G_{SII} measures were used to optimize the lengths L_1^{A2B} and L_2^{A2B} of the filter $\mathbf{w}_1^{A2B}$ and $\mathbf{w}_2^{A2B}$, respectively. The optimization of the filter lengths was carried out by varying the size of the first (or second) filter while the size of the second (or first) filter was kept fixed. Then, the polar diagrams of the omnidirectional microphone, the directional microphone and the optimized A2B were calculated, also, the transfer function of the speech and the noise signal was estimated between the directional microphone and the output of the adaptive beamformer for a specific angle of both sources (speech at 0^o and noise at 90^o).

3.3.2.2 Results and discussion

Figure 3.11 shows the average SNR in the noise reference of the A2B over all azimuth angles (between 0^o and 345^o). The SNR was roughly the same for $L_1^{A2B} > 6$ and increased for $L_1^{A2B} < 6$. Although a size of $L_1^{A2B} = 6$ is optimal for having a good cancellation of the speech signal in the noise reference, a size of $L_1^{A2B} = 10$ was chosen. This improves the definition of the frequency response of the filter and can result in a better robustness to the algorithm against a deviation of the characteristics of the microphones without increasing the cost of calculation significantly. **Figure 3.12** shows the average

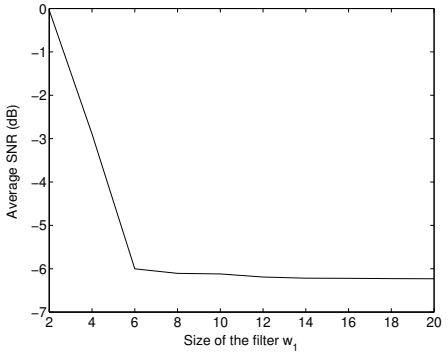


Figure 3.11: Influence of the filter size L_1^{A2B}, while the size of the second filter is kept fixed ($L_2^{A2B} = 30$).

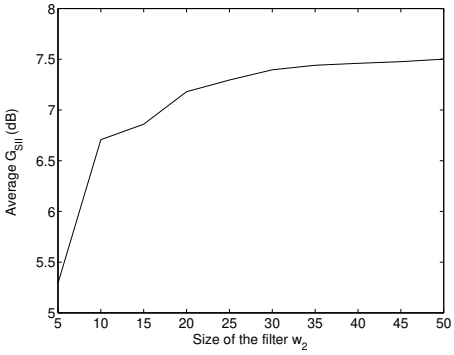


Figure 3.12: Influence of the filter size L_2^{A2B}, while the size of the first filter is kept fixed ($L_1^{A2B} = 10$).

G_{SII} between the omnidirectional microphone and the output of the A2B over all angles (between 0^o and 345^o). When $L_2^{A2B} > 30$, the G_{SII} stayed roughly the same, while for $L_2^{A2B} < 30$, the G_{SII} was observed to drop. For the real time implementation, $L_1^{A2B} = 10$ and $L_2^{A2B} = 30$ at a sampling frequency of 15.625kHz were chosen for the filters $\mathbf{w}_1^{A2B}$ and the $\mathbf{w}_2^{A2B}$, respectively.

In **figure 3.13**, data of the polar diagram are presented. The SNR_{SII} of the signals at the omnidirectional microphone, directional microphone and the output of the two-stage adaptive beamformer are shown as a function of the azimuth angle of the noise source. The SNR_{SII} of the omnidirectional microphone varied as a function of the angle, illustrating the head shadow effect of the dummy head. The SNR_{SII} at the output of the system was more than 11dB at all angles from 75^o to 270^o and was roughly 6dB or more between 45^o and 285^o. The improvement was mainly due to the adaptive processing for angles between 15^o and 175^o, and to the effect of the directional microphone and the head shadow for angles between 175^o and 345^o. Furthermore, the relatively broad angular response of the beamformer due to the small microphone spacing and the small filter length of $\mathbf{w}_1^{A2B}$ could be demonstrated with the simulation data in **figure 3.13**. Between 330^o and 45^o, which agreed with SNR-values of the output curve in **figure 3.13** below 3dB, everything could be considered as desired signal. Also in **figure 3.13** the SNR of the noise reference signal is shown for a single noise source as a function of angle relative to the direction of the speech source. Referring to the 0dB SNR input signals, the desired signals are clearly suppressed in the noise reference.

In **figure 3.14**, the transfer functions between the directional microphone (DO_{Hd}) and the output signals of the A2B for speech at 0^o and noise at 90^o are shown. The transfer function for the speech signal at 0^o was flat (this

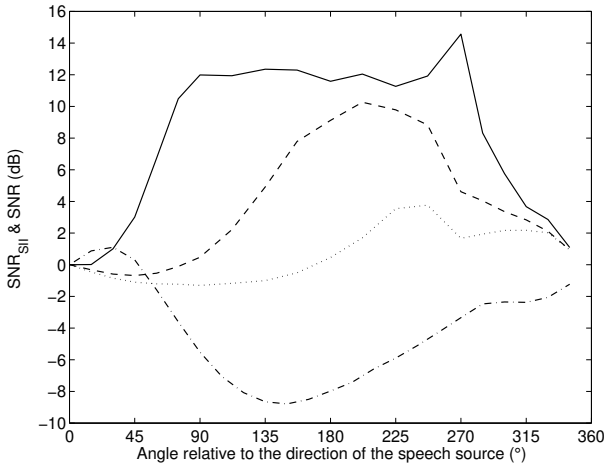


Figure 3.13: The four curves show the SNR (in dB) at the noise reference ($-\cdot$) and the SNR_{SII} (in dB) at the output of the two-stage adaptive beamformer ($-$), at the directional microphone ($--$) and at the omnidirectional microphone ($\cdots$). All as a function of the angle relative to the direction of the speech source.

means that the desired signal was not distorted during processing) and around 6dB. This amplification was due to the summation of the delayed version of the directional microphone and the filtered version of the omnidirectional microphone. The attenuation at the frequencies below 120Hz and above 7.5kHz was due to the DC filter and the anti-aliasing filter. An important attenuation of the noise source relative to the speech source was observed. This attenuation was frequency dependent: large at high frequencies and smaller at low frequencies, especially below 1kHz. In this experimental situation (noise coming from 90^o), the gain in speech intelligibility between the directional microphone and the output of the A2B was estimated to be 11.5dB (**figure 3.13**).

3.3.3 Perceptual evaluations

To validate the optimization and the physical evaluation of the two-stage adaptive beamformer, perceptual measurements were performed with five subjects, three normal hearing listeners and two hearing impaired patients.

3.3.3.1 Experimental set-up

The A2B strategy was implemented in a wearable research DSP-platform, called 'Audallion'. The Audallion platform was used before in other wearable

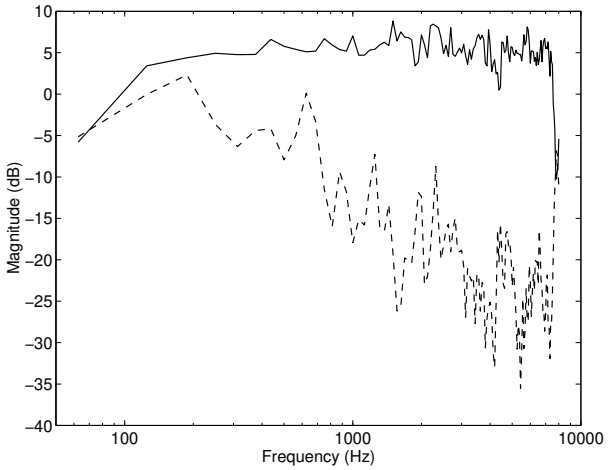


Figure 3.14: Magnitude (in dB) of the transfer function for speech-weighted noise at 0^o $(-)$ and speech-weighted noise at 90^o $(--)$ between the signals at the output of the directional microphone and at the output of the noise reduction algorithm.

DSP implementations, such as the Audallion BEAMformer® which was used for cochlear implant users. However, our beamformer strategy differed from the strategy used in the Audallion BEAMformer® (Schweitzer et al., 1996; Schweitzer, 1997; Figueiredo et al., 2001).

The Audallion was connected to the BTE hearing aids, and allowed the processing of two microphone signals with a DSP combined with a 16 bit ADC and DAC converter. The ADC sampling rate was 31.25 kHz, downsampled to 15.625 kHz. The DSP was a Motorola-56009 with a processor speed of 32MHz (16 MIPS). The size of the Audallion is 45x80x15mm and can easily be worn in a pocket. The listener could check the volume of the linear amplification and had the choice between three different signals: the omnidirectional microphone signal, the directional microphone signal and the output of the two-stage adaptive beamformer.

The experimental set-up used for the perceptual evaluation is shown in **figure 3.15**. The perceptual evaluations were carried out with the BTE hearing aid *DO* positioned on the dummy head in a test room ($\approx$25m^3 and average reverberation time $T_{60} \approx 0.35s$, the estimated direct-to-reverberant ratio was 4dB at one meter), which was also used for hearing aid fitting work. The two microphone signals were real-time processed in the Audallion. The output was routed to the neighbouring room where the signal was presented monaurally to the subjects through headphones TDH39 connected to an audiometer Madsen

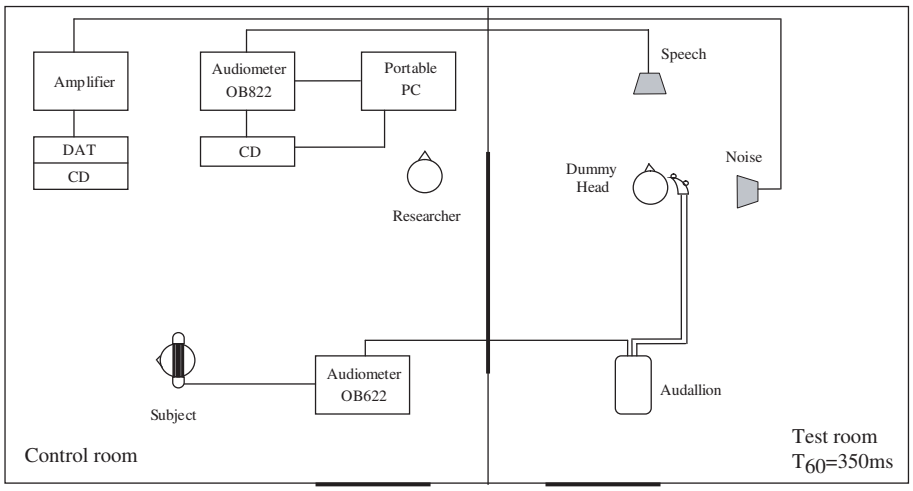


Figure 3.15: The experimental set-up for the perceptual evaluations of the two-stage adaptive beamformer.

OB622. The speech material and the competing noise were presented at 0^o and 90^o azimuth, respectively, relative to the hearing aid prototype positioned on the dummy head. The signals were produced by identical loudspeakers (Kenwood LS34) placed at a distance of one meter from the center of the head.

3.3.3.2 Subjects

Five subjects, three normal hearing and two hearing impaired, collaborated in this study. The normal hearing subjects had pure tone thresholds no higher than 15dBHL in the 250Hz to 8kHz range. The two hearing impaired subjects (subjects four and five) were experienced hearing aid users and had pure tone average (PTA) (1-2-4kHz) thresholds of 38 and 42dBHL at the tested (best) ear. Their impairment was purely sensorineural. Subjects four and five had a flat and a slopping (at about 15dB per octave) hearing loss, respectively. With the OB622 audiometer the overall level of the presented speech signals was increased to the most comfortable level for the hearing impaired subjects, to take into account their higher auditory thresholds. The frequency response was not compensated for the specific hearing loss.

3.3.3.3 Test materials and procedures

The SRT for three combinations of speech material and competing jammer sounds were measured twice for each of the two conditions: with and without the noise reduction processing active. In total, twelve SRT-values were obtained per subject (one test session of about forty minutes per subject). For the reference condition only the signal of the directional microphone DO_{Hd} was used. This already imposed a benefit against jammer signals from 90^o over an omnidirectional microphone of about 3dB in realistic acoustical environments (Hawkins and Yacullo, 1984; Leeuw and Dreschler, 1991; Wouters et al., 1999). Care was taken that the level of the speech was identical in both conditions, with or without the noise reduction. The sequence of the different test conditions and used lists of speech materials was changed in the experiment for every subject to counter eventual training and/or fatigue effects over the different conditions and materials.

The test materials consisted of sentences in speech-weighted steady noise, sentences in multitalker babble and two-syllable BLU words in speech-weighted steady noise (Plomp and Mimpen, 1979; Wouters et al., 1994). The adaptive procedure described in section 2.2.1.3 was used for the speech-in-noise evaluations. The SRT was obtained as the mean level of the speech at the ten and fifteen last responses to the presentation of thirteen sentences and twenty two-syllable words, respectively. These methods have shown to yield SRT-precisions within 1 and 1.3dB, respectively.

The multitalker babble was taken from the compact disc Auditory Tests (Revised) edited by Auditec of St. Louis. The steady speech weighted noise was specifically constructed for both speech materials separately. These materials were available on CD and were presented in the experimental set-up using two CD-players. The adaptive testing procedure was run automatically using a portable PC with infrared link that controls an OB822 audiometer (for the speech level changes) and CD-player (**figure 3.15**). The competing noises were presented at 70dBSPL.

3.3.3.4 Results

In **table 3.1**, the experimental results of the speech intelligibility tests in noise are expressed as speech-to-noise ratios in dB for 50% intelligibility. Subjects one to three were the normal hearing and subjects four and five were the hearing impaired subjects. With the directional microphone, the average SNR at 50% intelligibility for the sentences in speech weighted noise was $-3.3(1.2)$dB for the normal hearing subjects. The average SNR in multitalker noise was higher. For the hearing impaired both SNRs were higher than for the normal hearing, as expected. Comparing the A2B and DO_{Hd} data, SNR-improvements of 12.0(2.0)dB, 11.5(3.0)dB and 10.5(2.2)dB were obtained for

Table 3.1: The obtained SNR-values for 50% intelligibility (in dB) as measured at the output of the two-stage adaptive beamformer and at the directional microphone (the improvement was the difference between both) for the three combinations of speech and noise materials. Each condition was measured twice for five subjects. Subjects four and five were hearing impaired. The mean SNR-improvement and standard deviation for each combination is given.

	Sentences in speech weighted noise			Sentences in multitalker babble			Words in speech weighted noise		
Subject	Output	Directional	Improvement	Output	Directional	Improvement	Output	Directional	Improvement
1	-15.4	-3.8	11.6	-11.8	3.4	15.2	-16.4	-2.8	13.6
	-16.2	-1.8	14.4	-10.2	-0.2	10.0	-16.4	-7.6	8.8
2	-15.8	-2.6	13.2	-9.8	1.0	10.8	-16.0	-4.8	11.2
	-14.2	-2.6	11.6	-10.6	3.4	14.0	-13.4	-2.8	10.6
3	-14.2	-4.6	9.6	-8.2	3.0	11.2	-16.0	-4.8	11.2
	-13.8	-4.6	9.2	-8.6	-1.0	7.6	-18.4	-5.6	12.8
4	-11.4	-1.8	9.6	-7.0	4.6	11.6	-13.4	-3.4	10.0
	-13.0	0.6	13.6	-4.6	4.6	9.2	-12.2	-4.4	7.8
5	-11.8	0.2	12.0	-4.6	4.2	8.8	-10.0	-3.2	6.8
	-13.0	1.8	14.8	-5.4	11.4	16.8	-13.6	-1.6	12.0
Mean			12.0			11.5			10.5
STD			2.0			3.0			2.2

the sentences in speech-weighted noise, sentences in multitalker babble and two-syllable words in speech-weighted noise, respectively.

3.3.4 Discussion

An analysis of variance (ANOVA) with repeated measures on the within-subject factors processing (DO_{Hd} and A2B) and speech-in-noise combinations (sentences in speech-weighted noise, sentences in multitalker babble, two-syllable words in speech-weighted noise) was carried out using SPSS 9 software. The obtained improvements between DO_{Hd} and A2B were highly significant ($F=399$, $p<0.001$). There were significant differences between the obtained SNRs at 50% intelligibility for different speech-in-noise test materials ($F=168$, $p<0.001$, and the three test materials yield significantly different SNRs). However, the SNR-improvement between DO_{Hd} and A2B did not differ significantly for the different speech-in-noise combinations ($p=0.212$), nor between hearing impaired and normal hearing subjects ($p=0.736$), although the absolute SNR-values differed appreciably and significantly. The latter result needs to be considered with caution. Whereas the SNR-improvement was highly significant even with five subjects tested, the result of no difference between normal hearing and hearing impaired subjects was not representative because of the small number of subjects tested. This data demonstrated that similar SNR-improvements with

this noise reduction strategy could be obtained by hearing impaired subjects, even without compensating for the specific frequency-dependent hearing loss. The mean SNR-improvement of the dual microphone noise reduction strategy was 11.3dB (standard error 0.6dB) for one noise source (with stationary or speech-modulated temporal envelope) at 90^o relative to the speaker direction, relative to the effect of a directional microphone. This value agreed with the calculated result of 11.5dB intelligibility weighted SNR-improvement obtained in the physical evaluations. For normal hearing subjects, the SNR at 50% intelligibility for sentences in this multitalker babble was about 5dB higher than in stationary speech-weighted noise. With the gain of 11dB SRT reported here, a slightly reduced benefit due to multi-jammer-noise source environments or increased reverberation, should still allow an important improvement relative to the alternative of a single directional microphone configuration. In the next section, the adaptive beamformer is assessed in different jammer sound scenarios and compared to an omnidirectional microphone or a directional microphone for different types of hearing aids.

3.4 Large scale evaluation

In this section, a large-scale evaluation of the two-stage adaptive beamformer is presented. The performance for speech understanding in noisy environments of the A2B implemented in the Audallion is evaluated relative to a directional and an omnidirectional microphone, with two types of BTE hearing aids, *DO* and *OO* (**figures 3.2** and **3.3**). An extensive physical evaluation is performed to study the influence of the microphone configurations in the hearing aids on the performance of the adaptive beamformer. For the perceptual evaluation, the main perceptual research questions are related to the influence on the noise reduction performance of i) the spectro-temporal character of the jammer sound, ii) the jammer sound scene, iii) hearing impairment, iv) the basic microphone configuration in the hearing aid. To carry out this evaluation, four different speech materials are used, and also, three different jammer noise sounds, three jammer sound scenarios and two microphone configurations. Three groups of subjects take part in this study. The tests are carried out in moderately reverberant conditions (living room situation) and in highly reverberant conditions (diffuse noise). In total 1368 SRT-measurements are performed. The publications related to this section are (Maj et al., 2001b; Maj et al., 2003).

3.4.1 Experimental set-up

For the physical and the perceptual evaluation, the same test room was used. Since the latter had recently been installed, the reverberant properties

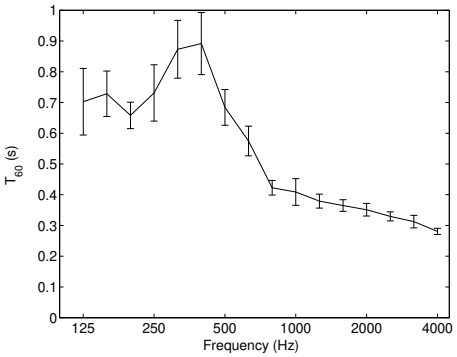


Figure 3.16: Reverberation time (T_{60}) in the reverberant room obtained for twelve different positions of a loudspeaker for an unmodulated speech weighted noise. The means and the standard deviations for sixteen of the one-third octave bands from 125Hz to 4kHz are represented. With the full-band speech weighted noise, an average of 0.76s is obtained for the T_{60}.

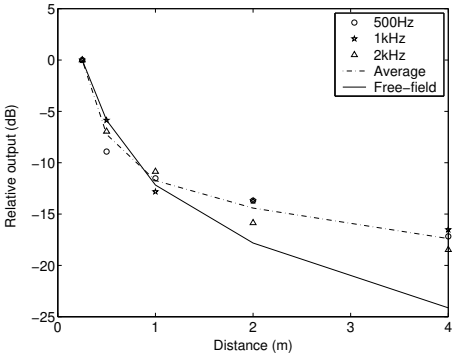


Figure 3.17: Theoretical free-field level of a sound source compared with the measured level in the test room for one single loudspeaker. The measurements were relative to the sound level at 25cm from the loudspeaker. The critical distance was about 2m over the average of 0.5, 1 and 2kHz.

of the chamber were characterized. This room had a volume of 70m^3 and the reverberation time (T_{60}) of the room was obtained for twelve different positions of a loudspeaker for an unmodulated speech-weighted noise. The mean of T_{60} for the twelve positions (and the standard deviation) is presented in **figure 3.16** for sixteen center frequencies from 125Hz to 4kHz of one-third octave bands. With the full-band speech weighted noise, an average T_{60} of 0.76s was obtained. The critical distance was measured with a sound level meter (Rion NA-27) in one-third octave bands. **Figure 3.17** shows the theoretical free-field level of a sound source compared with the measured level in the test room of one single loudspeaker. The measurements were relative to the sound level at 25cm from the loudspeaker. The critical distance was about two meters over the average of 0.5, 1 and 2kHz. A reference point in the room was defined where, for physical measurements, the center of the mannequin, the hearing aid in stand-alone configuration and, for perceptual measurements, the center of the head of the subject was positioned.

Figure 3.18 shows the experimental set-up used for the perceptual evaluation. In the test room, three loudspeakers (Yamaha CBX-S3) with approximately flat (±2dB for 150Hz-10kHz) frequency response were used for the speech and discrete noise sources: one for the speech signal in front of the subject at 0^o, and two for noise signals at 90^o (side of the hearing aid) and at

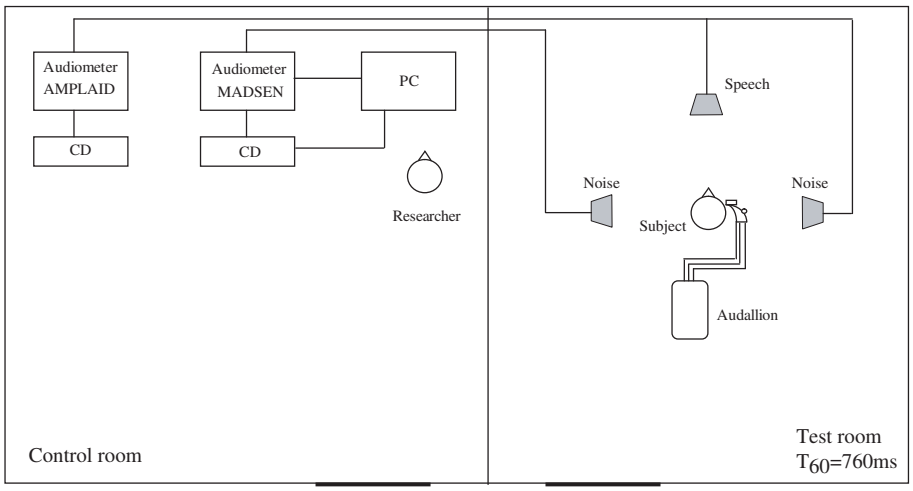


Figure 3.18: The experimental set-up for the large scale perceptual evaluations of the two-stage adaptive beamformer.

270^o (the side opposite to the hearing aid). The loudspeakers were calibrated separately to obtain the same sound level for each loudspeaker at the reference point (defined above). The middle of the loudspeakers was at the same height as the hearing aid of the subject (140cm) and the distance between each loudspeaker and the middle of the subject's head was one meter. For noise scenarios with more than one spectrally identical jammer noise, the presented sources were uncorrelated.

In the control room, a second room, the speech signals were sent out by a computer sound card (Sound Blaster 16 bits) and the noise signals by a SONY CD991 CD player. Afterwards, for amplification purposes and intensity control, the speech signal was sent through a MADSEN OB822 audiometer (channel right), the noise at 90^o through a MADSEN OB822 audiometer (channel left) and the noise at 270^o through an AMPLAID 309 audiometer (channel right). For the diffuse noise test, the approach of Veit and Sander (1987) was applied. Eight loudspeakers sent out an uncorrelated speech weighted noise signal and the sound level of each loudspeaker was adjusted to have the same power level (61dBSPL) at the reference point. With the eight loudspeakers together, a sound level of 70dBSPL was obtained. The reverberation of the test room was also increased to T_{60}=0.88s (for a speech-weighted noise) and the critical distance was approximately 0.8m. The same diffuse sound level ($\pm$0.5dB) was obtained in a cube of 50cm around the reference point. Compared to tests performed in anechoic or moderately reverberant conditions, tests carried out in diffuse noise are a kind of model of the most adverse listening environment.

For instance, the tests in moderately reverberant conditions and in a diffuse noise could refer to improvements obtained in a living room and in a very reverberant conference room, respectively.

3.4.2 Physical evaluations

The physical performance of the system was evaluated in two parts. The first part involved acoustic measurement in an anechoic chamber[3] with the hearing aids in stand-alone configuration. These measurements were used to calculate the DI and construct speech intelligibility weighted polar diagrams. The second part consisted of performing the evaluation in a reverberant chamber (described above) with the hearing aids mounted on a mannequin. Intelligibility weighted polar diagrams and transfer functions of speech and noise signals were calculated.

In the anechoic condition, the recordings were made with an ICRA-signal, an unmodulated noise representative of a male weighted idealized speech spectrum at normal effort (ICRA, 1997). In reverberant conditions, a stationary speech-weighted noise of the Dutch sentences spoken by a male speaker, the same used for the perceptual evaluation, was used for the recordings. In both conditions, the signals were sent through a loudspeaker located at one meter in the azimuth plane of the hearing aid (in stand-alone) or the center of the mannequin. The signals at the microphones of the hearing aids were recorded for different locations of the loudspeaker, corresponding to angles between 0^o (the endfire angle) and 360^o in steps of 15^o. The endfire angle 0^o was taken as the reference angle. The level at the loudspeaker was adjusted to get a sound level of 70dBSPL at the reference point.

The signals of the omnidirectional microphones and the directional microphones (software and hardware) were directly accessible. The signal of the two-stage adaptive beamformer changed with the adaptation of the second filter. To carry out the physical evaluation of the noise reduction algorithm, the coefficients were adapted on the noise signal until convergence and then they were kept fixed. Afterwards, the reference signal (at 0^o) and noise signals were filtered separately with the fixed coefficients. These two separately filtered signals were used for the physical evaluation of the two-stage adaptive beamformer.

3.4.3 Perceptual evaluations

The main research questions of this section were addressed using SRT-measurements of three different subject groups I, II and III consisting of ten normal hearing, five hearing impaired and seven normal hearing persons, re-

[3] Anechoic chamber at the physics department of the KULeuven.

Table 3.2: Overview of the different tests performed by three subject groups.

				Speech Material			
Group	Hearing aid	Noise config.	Noise material	Man / Sentences	Woman / Sentences	Man / Numbers	Woman / Numbers
I	*OO*	90°	SW	×	×	×	×
			ICRA	×	×	×	×
			Babble	×	×	×	×
	OO	90° / 270°	SW	×		×	
			ICRA				
			Babble	×		×	
II	*OO*	90°	SW	×	×	×	×
			ICRA	×	×	×	×
			Babble	×	×	×	×
	OO	90° / 270°	SW	×		×	
			ICRA				
			Babble	×		×	
III	*OO*	Diffuse	SW	×	×	×	×
			ICRA				
			Babble				
	DO	90°	SW	×	×	×	×
			ICRA				
			Babble				

spectively. **Table 3.2** gives an overview of the different tests that the three subject groups performed. The tests of the omnidirectional microphone, the directional microphone and the two-stage adaptive beamformer of hearing aids *OO* and *DO* were carried out with three spectro-temporally different jammer sounds (SW: unmodulated speech weighted noise, ICRA: modulated speech weighted noise, and Babble: multitalker babble) and in three different noise scenarios (single noise source at 90^o, noise sources at 90^o and 270^o relative to speaker position, diffuse noise scene). Four speech materials were used: sentences and numbers lists both from a male and a female speaker.

3.4.3.1 Subjects

Three groups of subjects participated in these measurements. Group I consisted of ten normal hearing listeners. Their ages ranged from twenty one to twenty six (mean of twenty two years) and their pure tone thresholds were less than or equal to 15dBHL in the 250Hz to 8kHz range.

Group II consisted of five hearing impaired subjects. They were experienced hearing aid users and had pure tone average (1-2-4kHz) thresholds of 60, 45, 35, 55, and 53dBHL at the tested ear with slopping hearing loss. They were selected from the patient population at the ENT departement of the KULeuven University Hospital. Their ages varied from twenty five to sixty two (mean of

fifty years). Their impairment was purely sensorineural. They could set frequency independent amplification with Audallion to a comfortable level, but for the SRT-measurements, the specific hearing loss was not compensated for in the amplification.

For group III, seven normal hearing listeners participated. Their ages ranged from twenty one to forty, with a mean of twenty seven years, and their pure tone thresholds were less than or equal to 15dBHL in the 250Hz to 8kHz range. Group III did not perform the same tests at the same time as group I. To avoid eventual fatigue and/or training effects in the data, two different groups of normal hearing listeners participated in this study. Also, all conditions could not be measured within the same group due to the limited availability of lists of the speech material (**table 3.2**).

The perceptual evaluation was carried out in monaural configuration and the non-test ear was acoustically blocked by an individualized deep and closed earmould. With the Audallion the overall level of the presented signals was adjusted to the most comfortable level. This level was kept fixed for all measurements. Care was taken that the level of the speech signal was identical for all conditions: omnidirectional microphone, directional microphone and output of the A2B.

3.4.4 Test materials

To measure the SRT, the adaptive procedure (section 2.2.1.3) was used and the level step size for each speech material was 2dB. In order to determine the SRT, the levels of the last seven and ten responses of a list of ten numbers and thirteen sentences, respectively, were averaged. The two noise sources were fitted separately, to get a constant sound level at the reference point. Thus, a sound level was obtained of 65dBSPL and 68dBSPL with one noise and two noise sources, respectively.

Sentences and numbers recorded by male and female voices were used as speech materials, and, unmodulated speech weighted noise (with the identical spectrum as the corresponding speech material), ICRA speech modulated noise and multitalker babble were used as noise materials.

The sentence speech materials are Dutch sentences developed by Versfeld et al. (2000). These sentences are an extension of the materials used by Plomp and Mimpen (1979) to measure speech reception thresholds. Two times thirty nine lists were available (thirty nine by a female talker and thirty nine by a male talker), each list contained thirteen sentences. The number speech material used in this study was recorded by four speakers (two male and two female), and consisted of forty different lists (ten per speaker) of ten different numbers in random order. Each list had one number from each decade in

every list. The recording and digital processing of this material were carried out identically to the development of the two-syllable word lists (Wouters et al., 1994). The number lists were evaluated in three steps by normal hearing subjects for intelligibility-in-noise tests. After each evaluation the level of each number was adjusted by a certain well-determined amount, dependent on the individual intelligibility score, to obtain an as similar as possible psychometric curve for all numbers. This was done with, in general, smaller adjustments relative to the fixed noise level in each step. Participants in this evaluation study were presented all numbers of the different speakers at three different sound levels. An analysis as described by Versfeld et al. (2000) was carried out for this homogenization procedure. The final result was that the SRT for these numbers one to one hundred in speech noise, weighted according to the spectrum of the numbers, is -10.5dB SNR with a slope around the fifty percent point of intelligibility of about fifteen percent per dB.

One of the three test noises was unmodulated noise speech weighted according to the spectrum of the specific speech materials used. The two other noises were modulated. The ICRA noise signal used in this work was a combination of six random Gaussian noises which were speech-shaped modulated according to the babble of six persons. The speech modulations consisted of signals of one female and one male talker, as well as two other female and two other male talkers each at -6dB relative to the first two talkers. The signals of all six speakers were three-band speech modulated noise, based on an idealized speech spectrum (Byrne et al., 1994) and produced at normal effort (ICRA, 1997). The multitalker babble was taken from the compact disk Auditory Tests (Revised) edited by Auditec of St. Louis.

3.4.5 Results and discussion

3.4.5.1 Physical evaluations

Figures 3.19 and **3.20** show the intelligibility weighted polar diagrams of both hearing aids *OO* and *DO* in stand-alone configuration and in anechoic conditions. As expected, the omnidirectional microphones (**figures 3.19** and **3.20**) had the same sensitivity for all angles and the software directional microphone (**figure 3.19**) had a hypercardioid polar diagram. The hardware directional microphone had a cardioid polar diagram. The polar diagrams of the output of the two-stage adaptive beamformer depended on the type of the hearing aid. The polar diagram with the hearing aid *DO* always presented a better noise reduction than the hearing aid *OO*. In **figures 3.19** and **3.20**, the patterns in the top-half and the bottom-half were symmetrical.

The *DI* of each signal was presented in **figure 3.21**. The *DI* of an omnidirectional microphone was expected to be 0dB for all frequencies. The measured *DIs* of the omnidirectional microphones were between about -1dB

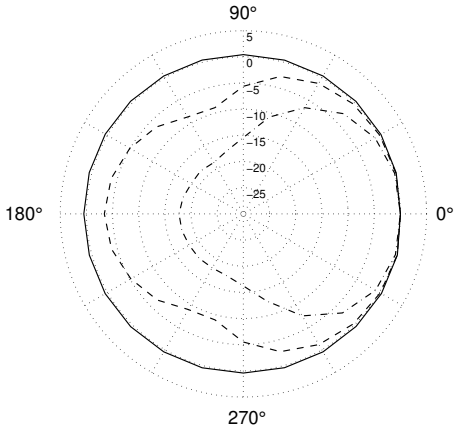


Figure 3.19: Intelligibility weighted polar diagrams (in dB) of the omnidirectional microphone (−), the software directional microphone (−−) and the adaptive beamformer (−·) of the hearing aid OO. The plotted values are relative to the angle of the direction of the speech source (0^o), when the SNR at the center of the head is 0dB, measured on the basis of speech noise in the stand-alone configuration and in anechoic conditions.

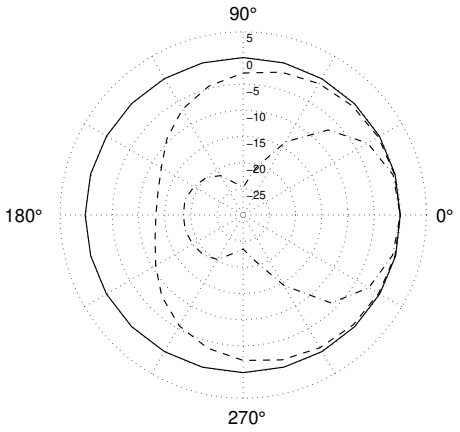


Figure 3.20: Intelligibility weighted polar diagrams (in dB) of the omnidirectional microphone (−), the hardware directional microphone (−−) and the adaptive beamformer (−·) of the hearing aid DO. The plotted values are relative to the angle of the direction of the speech source (^{o}), when the SNR at the center of the head is 0dB, measured on the basis of speech noise in the stand-alone configuration and in anechoic conditions.

and -0.5dB for the hearing aids OO and DO, respectively, and were roughly the same for all frequencies. These differences, between the expected and the measured values and also between the two omnidirectional microphones of both types of hearing aid (OO and DO), were mainly due to the position of the devices. Certainly, the hearing aids were not positioned on the horizontal plane and at the same place for both measurements. However, these differences were small and the DI_{AI} equals -1dB and -0.2dB for the hearing aids OO and DO, respectively.

The DI of the hardware directional microphone was about 4dB at low frequencies (160Hz-1.6kHz) and decreased slightly at higher frequencies. The software directional microphone had low values of the DI at low frequencies (below 800Hz). Thompson (1999) argues that it is necessary that the sensitivity of the two omnidirectional microphones of the hearing aid be adequately matched in both magnitude and phase to use the software directional microphone. Otherwise, this affects the performance of the DI at low frequencies. Above 1kHz, the software directional microphone was more directional than

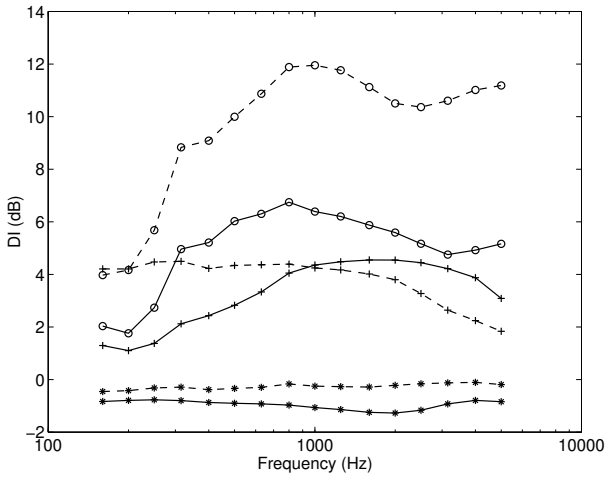


Figure 3.21: The Directivity Index DI (in dB), of both hearing aids OO $(-)$ and DO $(--)$, of the omnidirectional microphones $(*)$, the directional microphones $(+)$ and the output of the two-stage adaptive beamformer (o) are presented. The plotted values are relative to the angle of the direction of the speech source (0^o), when the SNR at the center of the head is 0dB, measured on the basis of speech noise.

the hardware directional microphone. The DI_{AI} of the software directional microphone was 3.8dB and the DI_{AI} of the hardware directional microphone was 3.7dB. The DI of the output signals of the two-stage adaptive beamformer for both hearing aids showed a similar pattern of performance, although the levels differed. The DI_{AI} of the output with hearing aid OO was 5.6dB and with the hearing aid DO was 10.7dB.

The intelligibility weighted polar diagrams, in reverberant conditions, of the omnidirectional microphones of the hearing aids OO and DO on the mannequin (**figures 3.22** and **3.23**) showed higher sensitivity for the angles 0^o to 180^o in comparison with other angles. These effects were mainly due to the head shadow effect. Compared to anechoic stand-alone conditions (**figures 3.19** and **3.20**), the performance of the software directional microphone (**figure 3.22**), the hardware directional microphone (**figure 3.23**) and the output signal of the two-stage adaptive beamformer for both types of hearing aids decreased and the polar dependencies changed appreciably. For hearing aid OO (**figure 3.22**) and DO (**figure 3.23**) the directional microphones always performed better than the omnidirectional microphones. The beamformer technique, for both hearing aids, outperforms (or provided equal performance for few angles) the

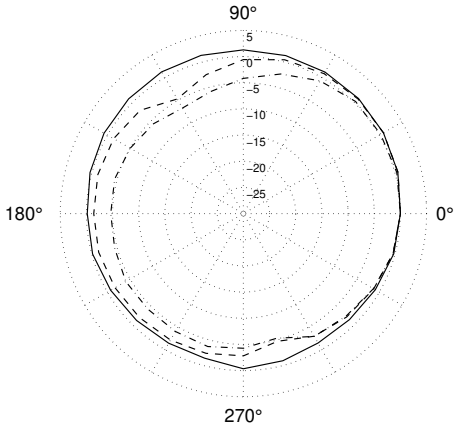


Figure 3.22: Intelligibility weighted polar diagrams (in dB) of the omnidirectional microphone (−), the software directional microphone (−−) and the adaptive beamformer (−·) of the hearing aid OO. The plotted values are relative to the angle of the direction of the speech source (0^o), when the SNR at the center of the head is 0dB, measured on the basis of speech noise on the mannequin and in reverberant conditions (T_{60}=0.76s).

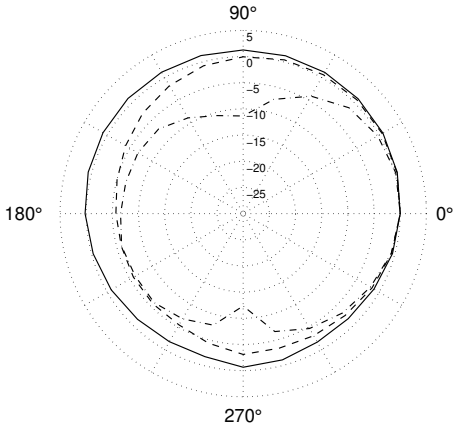


Figure 3.23: Intelligibility weighted polar diagrams (in dB) of the omnidirectional microphone (−), the hardware directional microphone (−−) and the adaptive beamformer (−·) of the hearing aid DO. The plotted values are relative to the angle of the direction of the speech source (^{o}), when the SNR at the center of the head is 0dB, measured on the basis of speech noise on the mannequin and in reverberant conditions (T_{60}=0.76s).

directional microphones. The difference in performance between the hardware directional microphone and the software directional microphone depended on the location of the noise source. The best performance was obtained with the output signal of the two-stage beamformer with hearing aid DO.

In **figure 3.24**, the transfer functions between the directional microphones (software and hardware) and the output of the two-stage adaptive beamformer, for speech at 0^o and noise at 90^o, are presented. On the one hand, the transfer function for the speech signal at 0^o was approximately flat and about 0dB for both hearing aids. This confirms again that the desired signal was not distorted by the adaptive beamformer. On the other hand, there was an important attenuation of the noise signals relative to the speech signals. However, this attenuation of the noise signal was more significant when the two-stage adaptive beamformer was used with the hearing aid DO for the frequencies 160-2.5kHz, which are crucial for speech understanding. The intelligibility weighted gain in SNR, G_{SII} was calculated for this sound scenario (speech at 0^o and noise at 90^o). Relative to the omnidirectional microphone,

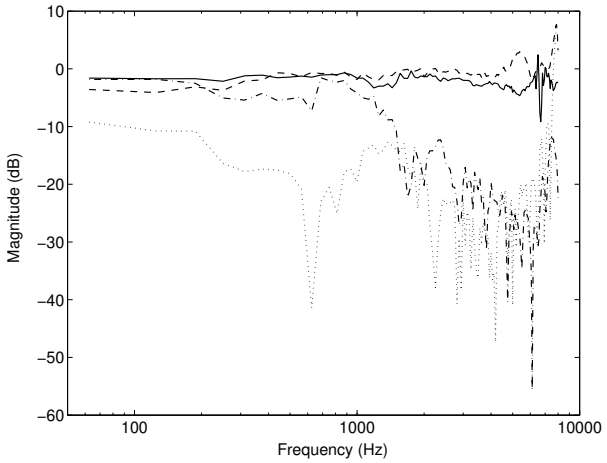


Figure 3.24: Magnitude (in dB) of the transfer function for speech-weighted noise at 0^o, $(-)$ for the hearing aid OO and $(--)$ for the hearing aid DO, and speech-weighted noise at an angle of 90^o, $(-\cdot)$ for the hearing aid OO and $(\cdots)$ for the hearing aid DO. These transfer function are between the software directional microphone and the output of the two-stage adaptive beamformer for the hearing aid OO, and between the hardware directional microphone and the output of the two-stage adaptive beamformer for the hearing aid DO.

$G_{SII}s$ of 2.4dB and 6.8dB were obtained for the software directional microphone and the adaptive beamformer with the hearing aid OO, respectively. The G_{SII} equaled 1.3dB for the hardware directional microphone and 11.4dB for the noise reduction algorithm relative to the omnidirectional microphone.

3.4.5.2 Perceptual evaluations

Tables 3.3 and **3.4** show the improvements (in dB) of the SRT relative to the omnidirectional microphone for the two-stage adaptive beamformer and the directional microphone for both types of hearing aids OO and DO. Every data cell corresponds to the mean (and the standard deviation) of all subjects in one group in each condition of the tests as defined in **table 3.2**. **Table 3.3** shows the SRT improvements between the omnidirectional microphone and the two-stage adaptive beamformer and **table 3.4** shows the SRT improvements between the omnidirectional microphone and the directional microphone. This is presented for the four different speech materials (sentences and numbers lists, spoken by male and female speakers) and the three noise materials (SW, ICRA

and Babble), the jammer sound scenes (90^o, $90^o/270^o$ and diffuse) and the three groups of subjects (I, II and III). The mean (and the standard deviation) of all values for each jammer sound scene of each group of subjects is given in column nine of **tables 3.3** and **3.4** (SRT-improvements between the omnidirectional microphone and the noise reduction algorithms, the directional microphone and the output of the A2B).

3.4.5.2.1 Spectro-temporal character of the jammer sound

The SRT-improvements depended on the jammer sound scenario and the type of the hearing aid. Indeed, the more complex of the jammer sound scene, the lower the SRT-improvements.

An analysis of variance (ANOVA) for repeated measures was carried out using SPSS 10.0 software. For the SRT-measurements of group I and group II using the hearing aid *OO* with a jammer sound at 90^o, the tests on the within-subject effects showed that there were significant SRT differences among the speech materials ($F=20.4$, $p<0.001$), and between the different noises ($F=98.7$, $p<0.001$). The interaction between the speech materials and the noises was also significant ($F=7.8$, $p<0.001$). From tests of within-subject contrasts, the difference between the results for the speech-weighted noise and ICRA noise was significant ($F=15.7$, $p<0.05$) and was highly significant with multitalker babble ($F=119.8$, $p<0.001$). Relative to the omnidirectional microphone, significant SRT differences were found with the software directional microphone ($F=90.2$, $p<0.001$) and the beamformer technique ($F=322.3$, $p<0.001$). Moreover, the analysis showed that these differences were improvements independent of the specific speech and noise materials used because the interaction terms for the speech and the noise materials used were not significantly different. The results indicate that the three jammer sounds and the four speech materials did not have an influence on noise reduction performance for the normal hearing and hearing impaired subjects (group I and II).

For a single jammer sound source at 90^o, the SNR-improvement between the omnidirectional microphone and the two-stage adaptive beamformer was 7.8dB for the normal hearing and 7.5dB for the hearing impaired listeners. These results can be compared with a recent study (Ricketts and Henry, 2002). Ricketts and Henry evaluated an adaptive directional microphone for hearing aids. In a less reverberant room ($T_{60} \approx 400 - 450ms$), with two noise sources at 75^o and 105^o, an SRT-improvement of about 2dB was obtained relative to an omnidirectional microphone. For the software directional microphone, the SNR-improvement was 3.1dB and 4.6dB for the normal hearing and the hearing impaired groups respectively. This improvement obtained with the software directional microphone corresponds to the values measured in other studies (Hawkins and Yacullo, 1984; Leeuw and Dreschler, 1991; Wouters et al., 1999).

Table 3.3: The average differences in SRT (in dB) for each group of subjects between the two-stage adaptive beamformer compared to the omnidirectional microphone for the different conditions. The means (and the standard deviation) of all values for different speech and noise test materials are given in column nine.

				Omnidirectional microphone-Two-stage adaptive beamformer				
	Hearing aid	Noise config.	Noise material	Male / Sentence	Female / Sentence	Male / Number	Female / Number	Mean
Group I	OO	90°	SW	8.7 (2.8)	7.6 (1.9)	7.2 (1.9)	8.2 (1.9)	7.8 (2.4)
			ICRA	8.0 (2.0)	9.6 (2.5)	5.5 (2.4)	8.1 (2.8)	
			Babble	8.8 (1.8)	7.7 (1.9)	7.7 (2.2)	7.0 (3.4)	
	OO	90° / 270°	SW	6.7 (2.5)	-	6.5 (3.1)	-	6.1 (3.0)
			ICRA	-	-	-	-	
			Babble	6.4 (3.0)	-	4.9 (3.4)	-	
Group II	OO	90°	SW	7.1 (2.9)	8.2 (4.0)	5.8 (2.8)	6.3 (2.1)	7.5 (3.4)
			ICRA	8.8 (2.8)	6.2 (4.1)	6.4 (1.8)	8.6 (4.2)	
			Babble	8.4 (3.4)	9.0 (4.6)	7.4 (1.2)	7.4 (5.8)	
	OO	90° / 270°	SW	6.2 (2.8)	-	6.0 (1.9)	-	5.8 (2.3)
			ICRA	-	-	-	-	
			Babble	5.8 (2.9)	-	5.2 (2.1)	-	
Group III	OO	Diffuse	SW	2.5 (1.6)	2.8 (2.0)	3.8 (1.5)	1.2 (2.3)	2.6 (2.0)
	DO	90°	SW	12.2 (1.5)	11.0 (2.3)	11.7 (3.3)	10.7 (2.4)	11.4 (2.4)

Table 3.4: The average differences in SRT (in dB) for each group of subjects for the directional microphone compared to the omnidirectional microphone for the different conditions. The means (and the standard deviation) of all values for different speech and noise test materials are given in column nine.

				Omnidirectional microphone-Directional microphone				
	Hearing aid	Noise config.	Noise material	Male / Sentence	Female / Sentence	Male / Number	Female / Number	Mean
Group I	OO	90°	SW	4.4 (1.8)	3.7 (2.1)	1.5 (1.8)	2.5 (1.7)	3.2 (2.4)
			ICRA	2.6 (3.1)	5.2 (2.4)	2.1 (2.5)	3.1 (1.6)	
			Babble	5.0 (2.6)	2.7 (1.8)	2.1 (2.6)	3.0 (2.6)	
	OO	90° / 270°	SW	4.7 (2.2)	-	2.9 (2.4)	-	3.5 (2.2)
			ICRA	-	-	-	-	
			Babble	4.0 (2.0)	-	2.4 (1.5)	-	
Group II	OO	90°	SW	4.6 (1.8)	5.3 (2.5)	2.2 (2.8)	4.0 (3.7)	4.6 (3.0)
			ICRA	5.0 (3.9)	4.6 (4.0)	4.3 (1.0)	4.5 (3.8)	
			Babble	5.4 (3.5)	5.3 (3.9)	5.4 (3.2)	4.4 (3.0)	
	OO	90° / 270°	SW	3.0 (2.9)	-	2.9 (2.8)	-	3.3 (2.4)
			ICRA	-	-	-	-	
			Babble	4.2 (1.8)	-	3.2 (2.4)	-	
Group III	OO	Diffuse	SW	1.5 (1.2)	2.4 (1.4)	2.3 (1.9)	1.1 (2.4)	1.8 (1.8)
	DO	90°	SW	3.3 (1.5)	3.7 (1.8)	3.0 (2.1)	1.7 (2.7)	2.9 (2.1)

For normal hearing subjects (group I), the SNR at 50% intelligibility for the four speech materials was about 0.2dB for the ICRA noise and 4.1dB for the multitalker babble, higher than in stationary speech-weighted noise. For hearing impaired subjects (group II), in comparison with the speech-weighted noise, the SNR at 50% intelligibility for the four speech materials in ICRA noise and multitalker babble was about 2dB and 4.4dB higher, respectively.

With group I, a second SRT measurement (a few weeks after the first measurement) was performed with the hearing aid OO and the twelve different speech-noise material combinations with a jammer sound at 90^o. From a separate ANOVA analysis, it was shown that the effect between the first and the second measurements were not significantly different.

3.4.5.2.2 Hearing impairment

Using the data obtained for SRT-measurements as described above, statistical analysis showed that the difference between the omnidirectional microphone and the adaptive beamformer for the normal hearing and hearing-impaired listeners was not significant (F=0.16, p=0.688). Thus, the results of the two groups (I and II) could be averaged. The resulting SNR-improvement was 7.7dB. The SNR-improvement between the omnidirectional microphone and the software directional microphone was 3.2dB for the normal hearing listeners and 4.6dB for the hearing impaired listeners. This difference was not significant (F=3.0, p=0.104). The SNR-improvement averaged over the two groups was 3.9dB. The latter result needs to be considered with some caution. Whereas the SNR-improvement was significant with fifteen subjects tested (ten normal hearing and five hearing impaired listeners) the result that there was no difference between normal hearing and hearing impaired subjects may not be representative because of the small number of hearing impaired subjects. However, this data demonstrated that similar SNR-improvements could be obtained with the directional microphone and the two-stage adaptive beamformer by hearing impaired subjects, even without compensating amplification for the specific frequency-dependent hearing loss.

3.4.5.2.3 Jammer sound scene

With one jammer sound at 90^o and with hearing aid OO, the gain in SRTs was 7.8dB for group I and 7.5dB for group II when the adaptive beamformer was used. With the software directional microphone, the improvement was 3.1dB for group I and 4.6dB for group II. With two uncorrelated jammer sounds at 90^o and 270^o, the performance obtained with the beamformer technique decreased to 6.1dB for group I, and, 5.8dB for group II. With the software directional microphone, remarkably, performance improved for group I when the jammer sound sources increased from one to two (3.1dB and 3.5dB respectively), but decreased for group II, 4.6dB with one jammer source and 3.3dB for two jammer sound sources. In a diffuse noise case, the software di-

rectional microphone resulted in an improvement of 1.8dB and the adaptive beamformer resulted in an improvement of 2.6dB.

Statistical analyses were also carried out for the different jammer sound scenarios (90^o, $90^o/270^o$ and diffuse noise) and subject groups I, II and III. Significant differences were found among the three jammer sound scenarios ($F=8.8$, $p<0.001$). The jammer sound scenario with a single noise source at 90^o differed from the diffuse noise ($F=1.38$, $p<0.001$). The difference between the single jammer at 90^o and the double jammer at 90^o and 270^o was significant but the effect was small ($F=0.324$, $p=0.036$).

Figure 3.22 shows that the physical SNR-improvement at the output of the adaptive beamformer with jammers at 90^o or 270^o was roughly the same ($\approx$5dB). However, for the omnidirectional microphone the noise source at 270^o was attenuated (SNR>0) and the noise source at 90^o was amplified (SNR<0). This was due to the head shadow effect. At 270^o, the attenuation of the jammer sound was mainly brought about by the directional microphone, and at 90^o, by the processing of the adaptive beamformer. The statistical analysis showed that there was a small difference between one and two jammer sound scenarios. This means that the noise source at 270^o was attenuated enough by the head shadow and the pattern of the directional microphone, and that the processing of the adaptive beamformer mainly focused on the noise source at 90^o.

In diffuse noise, the performance of the adaptive beamformer decreased a lot. The SNR-improvement of the adaptive technique was only 2.6dB relative to the omnidirectional microphone signal and 0.8dB relative to the software directional microphone. The noise reduction performance of the adaptive beamformer depended on the correlation of the noise signal between the two microphones. The higher the level of reverberation, the lower the correlation between the signals of the two microphones and the worse the performance in noise reduction. Also, the performance of the first stage of the two-stage adaptive beamformer, which was fitted in the anechoic condition, failed in highly reverberant conditions. The cardioid-like spatial characteristic, obtained at the noise reference in anechoic condition (cardioid focused to the back hemisphere, null at 0^o), became an omnidirectional spatial characteristic in highly reverberant conditions.

Greenberg and Zurek (1992) argued that the weights of the adaptive filter tend to zero when the reverberations are strong. The SNR-improvement between the omnidirectional and the software directional microphone (1.8dB) or between the omnidirectional microphone and the adaptive beamformer (2.6dB) was small but still important for hearing-aid users. In critical listening conditions (close to 50% of speech understood by the listener) an improvement of 1dB in SNR corresponds to an increase in speech understanding of about 15% in every day speech communication (Plomp and Mimpen, 1979).

A good noise reduction strategy should have adaptive processing (adaptive beamformer) for low reverberations and a fixed processing (software or hardware directional microphone configuration) for high reverberation. In this way, in different acoustical situations of low and high reverberation, optimal noise reduction can be obtained at the output of the system.

3.4.5.2.4 Microphone configuration

From columns nine (**tables 3.3** and **3.4**), it is seen that the adaptive beamformer always outperformed the directional microphone. The software directional microphone of *OO* had roughly the same performance as the hardware directional microphone of *DO* (3.1dB and 2.9dB respectively). However, when the output of the adaptive beamformer was used, there was a clear benefit by using hearing aid *DO* (11.4dB) compared to hearing aid *OO* (7.8dB).

The hearing aid types *OO* and *DO* were statically analyzed in more detail with subject group I and III, the jammer sound scenario at 90^o and for the four different speech materials. The difference between the software directional microphone and the hardware directional microphone was not significant. However, the results obtained with the adaptive beamformer were significantly different for the hearing aid type *OO* or *DO* ($F=19.8$, $p<0.001$). Relative to the omnidirectional microphone, with the hearing aid type *OO*, the SNR-improvement was 3.1dB at the software directional microphone and 7.8dB at the output of the adaptive beamformer. With the hearing aid *DO*, the SNR-improvement was 2.9dB and 11.4dB, respectively.

The configuration of the microphones had an influence on the performance of the adaptive beamformer. The beamformer with hearing aid *DO* resulted in an additional SRT improvement of 3.6dB in comparison with hearing aid *OO*. The performance of the adaptive beamformer seems to be dependent on the number of independent microphone entry ports, i.e. the number of degrees of freedom, which is three for hearing aid *DO* and two for hearing aid *OO*. The more independent sensor inputs, the better and the more independent jammer sound sources can be nulled out with appropriate signal processing.

The difference between the *OO* and *DO* hearing aids was already predicted by the acoustic evaluation. Indeed, the DI_{AI} values were the same for the directional microphones of *OO* and *DO* (software 3.8dB and hardware 3.7dB). But at the output of the adaptive beamformer, the DI_{AI} with the hearing aid *DO* (10.7dB) was approximately twice the DI_{AI} with *OO* (5.6dB). Also, as shown in recent studies (Ricketts, 2000a; Ricketts, 2000b), the directivity index and polar diagram can be strongly linked to the improvement of the speech intelligibility in noise.

The data obtained with hearing aid *DO* can be compared to other studies in the literature and also to our previous studies (Wouters and Vanden Berghe,

2001; Wouters et al., 2002) where a similar microphone configuration and adaptive beamformer was used. In the first study (Wouters and Vanden Berghe, 2001), the perceptual tests were performed with cochlear implant patients in a moderately reverberant condition, the same as in the present study, and a gain in speech intelligibility of about 10dB was obtained with the adaptive beamformer relative to the hardware directional microphone. In the second study (Wouters et al., 2002), the perceptual measurements were carried out in a less reverberant room ($T_{60} \approx 0.45ms$) with normal hearing listeners and hearing impaired subjects. Relative to the hardware directional microphone, an improvement of about 11dB was obtained with the adaptive beamformer. In these two latter studies, the hearing aid was worn by a mannequin during the test instead of real life positioning of the hearing aid on the subject as in this study. The comparison between the tests carried out with the hearing aid DO in this study and the previous studies shows that the speech intelligibility performed with a mannequin overestimated detrimental and also that reverberation has an important effect on performance.

Furthermore, the performance metric G_{SII} tended to underestimate the results obtained with the perceptual evaluation. With hearing aid OO, G_{SII}s of 2.4dB and 6.8dB were obtained for the software directional microphone and the adaptive beamformer respectively, compared to 3.1dB and 7.8dB with the perceptual evaluation. With hearing aid DO, the G_{SII} (1.3dB) also underestimated the SNR-improvement of the hardware directional microphone (2.9dB) but predicted the SNR-improvement obtained at the output of the adaptive beamformer (11.4dB).

3.4.6 Summary

The answers to the four research questions presented in the introduction of this section are summarized as follow, i) There was no influence of the spectro-temporal character of the jammer sound as well as the speech sounds on the noise reduction performance of the adaptive beamformer. The analysis of the SRT for the four different speech materials and three different jammer sounds demonstrated that the obtained improvements in SRT were independent of these specific speech-noise combinations, although the absolute SRT-level might differ widely. ii) Between the two jammers noise scenario, namely 90^o and $90^o/270^o$, a small difference was found. The noise scenario $90^o/270^o$ was found to be inappropriate for studying the effect of the adaptive beamformer against two noise sources. The noise source at 270^o was mainly attenuated by the head shadow and the directional microphone. In diffuse noise, only small improvements were obtained. iii) There was no influence of hearing impairment on the noise reduction performance according to our data. It was shown that the speech recognition of hearing impaired listeners (PTA≈40-60dBHL) benefited equally well from this speech-in-noise enhancement, even when a frequency-independent-amplification was applied. iv) There was an

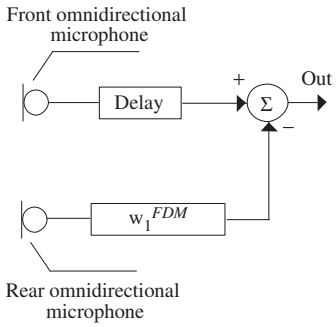


Figure 3.25: Representation of the fixed directional microphone (FDM). The FDM signal is the difference between the front microphone signal and a delayed version of the rear microphone signal. The delay is implemented by a filter operation (w_1^{FDM}).

influence of the basic microphone configuration on the noise reduction performance of the two-stage adaptive beamformer. With a noise source at 90^o, a SRT-improvement of about 7.8dB was obtained with the hearing aid OO versus 11.4dB with the hearing aid DO.

3.5 Comparison of the A2B with an adaptive directional microphone

In the present section, the two-stage adaptive beamformer is implemented in a commercial hearing aid and a comparison with a fixed directional microphone and an adaptive directional microphone (Luo et al., 2002; Ricketts and Henry, 2002) is performed. The device is a commercial GNReSound Canta7 BTE hearing aid with two omnidirectional microphones and was chosen because it contains a programmable DSP. In the two-microphone configurations, the acoustical characteristics (magnitude and phase) differ for each microphone and this may have an impact on the performance of the noise reduction algorithms. To compensate for these differences, the physical evaluation demonstrates the importance of an individual calibration procedure of the fixed filters. Then, polar diagrams and directivity indices are evaluated in anechoic and/or reverberant acoustical conditions. The perceptual evaluation is carried out in reverberant conditions by fifteen normal hearing subjects in different

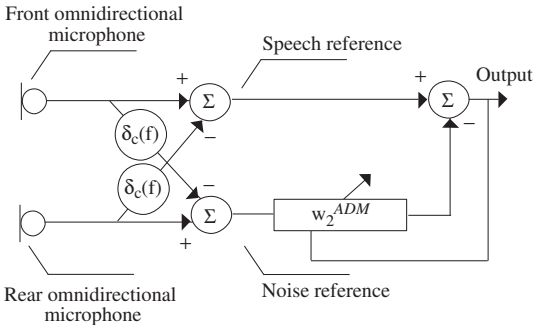


Figure 3.26: Representation of the adaptive directional microphone (ADM) where the signals of the speech reference and the noise reference are created with two software directional microphones. The delay operation $\delta_c(f)$ allows to obtain a front cardioid (null at 180^o) and a rear cardioid (null at 0^o). The filter of the ANC has one coefficient and is updated during processing.

jammer sound source scenarios. The publications related to this section are (Maj et al., 2003a; Maj et al., 2003b).

3.5.1 Noise reduction algorithms

3.5.1.1 Fixed directional microphone

The fixed directional microphone (FDM) signal is obtained as the difference between the signal of the front omnidirectional microphone and a delayed version of the rear microphone signal (**figure 3.25**). The delay operation is implemented by a filter operation ($\mathbf{w}_1^{FDM}$), which in our case has ten filter taps. The filter coefficients are optimized to obtain a hypercardioid polar diagram in anechoic conditions. This strategy differs from the conventional software directional microphone (Thompson, 1999), which has only one filter tap.

3.5.1.2 Adaptive directional microphone

The adaptive directional microphone (ADM) is illustrated in **figure 3.26** and is similar to state-of-the-art solutions in modern commercial digital hearing aids, such as the Phonak Claro and GNReSound Canta7, (Luo et al., 2002; Ricketts and Henry, 2002). Two software directional microphones create reference signals, namely the speech reference and the noise reference. The parameters of the software directional microphone are chosen to give a cardioid spatial

characteristic in anechoic conditions. The speech reference and the noise reference are made with a front cardioid (e.g. null at 180^o) and a rear cardioid (e.g. null at 0^o), respectively. The signals of the software directional microphones, the speech reference and the noise reference, are connected to an ANC. The filter w_2^{ADM} of the ANC in our case has one tap and can be updated by means of classical adaptive algorithms (Haykin, 1996). Also, a constraint is applied on the coefficient of the ANC. This constraint allows the adaptation of the coefficient when a source is at the back hemisphere (e.g. 90^o-270^o where the source is considered as noise) and stops the adaptation when a source is at the front hemisphere (where the source is considered as speech). This avoids the cancellation of the speech signal at the output of the ANC.

3.5.2 Experimental set-up

The noise reduction algorithms were implemented in the GNReSound Canta7 hearing aid and the output signals of the different algorithms as well as the front omnidirectional microphone signal were available at the loudspeaker of the hearing aids based on a selection program. For both evaluations, the choice of the program and parameter settings were done from a PC-platform, with software from Audiologic and GNReSound, through an interface device (HI-PRO). For the physical and the perceptual evaluation in reverberant conditions, the same test room as for the large scale evaluation of the A2B was used (section 3.4.1). All loudspeakers were positioned at one meter from the center of the head of the mannequin, for the physical measurements, and from the center of the head of the subjects for the perceptual measurements.

3.5.2.1 Hearing aid

The noise reduction algorithms were implemented for behind-the-ear (BTE) GNReSound Canta7 hearing aids (**figure 3.27**). These hearing aids are a powerful programmable digital platform. Two omnidirectional microphones (Microtronic-9667GX1) are mounted in an endfire array configuration spaced 1.6cm apart. For these tests, linear amplification was used in the hearing aid and the systems for compression or feedback control were switched off. Five Canta7 hearing aids, as obtained from the fabrication line, were used in this study.

3.5.2.2 Physical evaluation

For the physical evaluation, the loudspeaker of the hearing aid was connected to an ear simulator coupler (Bruel & Kjaer 4157) mounted on a microphone with a preamplifier (Bruel & Kjaer 4165 and Bruel & Kjaer 2639). The signal of this microphone was amplified by a Bruel & Kjaer (2610) amplifier and was digitized, at a sampling frequency of 16kHz, with a PC-platform us-

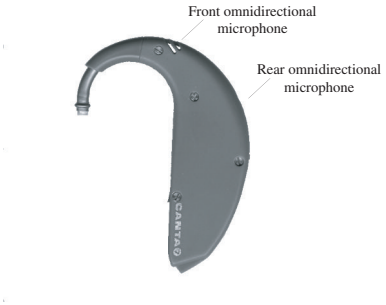


Figure 3.27: GNReSound Canta7 hearing aid where two omnidirectional microphones are mounted in an endfire array configuration spaced 1.6cm apart.

ing a LynxOne sound card with 16-bit analog-to-digital conversion. The test signals were sent through a loudspeaker located at one meter in the azimuth plane of the hearing aid (in stand-alone) or from the center of the head of the mannequin. The signals at the loudspeaker of the hearing aids were recorded for different locations of the loudspeaker, corresponding to angles between 0^o (endfire angle) and 360^o in steps of 15^o. The level at the loudspeaker was adjusted to obtain a sound level of 65dBSPL at the reference point, with the head 'in absence'. The physical measurements were carried out in an anechoic[4] and in a reverberant test room. The latter was also used for the perceptual evaluation.

3.5.2.3 Perceptual evaluation

For the perceptual evaluation, five loudspeakers (Yamaha CBX-S3) were used for the speech and discrete noise sources in the test room: one for the speech signal in front of the subject at 0^o, and four for presenting noise signals at 45^o, 90^o (side of the hearing aid), 180^o and 270^o (**figure 3.28**). The loudspeakers were calibrated separately to obtain the same sound level at the reference point for each loudspeaker (section 3.4.1). The middle of the loudspeakers was at the same height as the reference point (140cm) and the distance between each loudspeaker and the middle of the subject's head was one meter. For noise scenarios with more than one spectrally identical jammer noise, the presented signals were uncorrelated. In the control room, the speech signals were output by a computer sound card (Sound Blaster 16 bits) and the noise signals by SONY CD991 CD players. For amplification purposes and intensity control, the speech signal was sent through a MADSEN OB822 audiometer (channel right), the noises at 45^o or 90^o (depending on the noise scenario)

[4] Anechoic chamber at the physics department of the KULeuven.

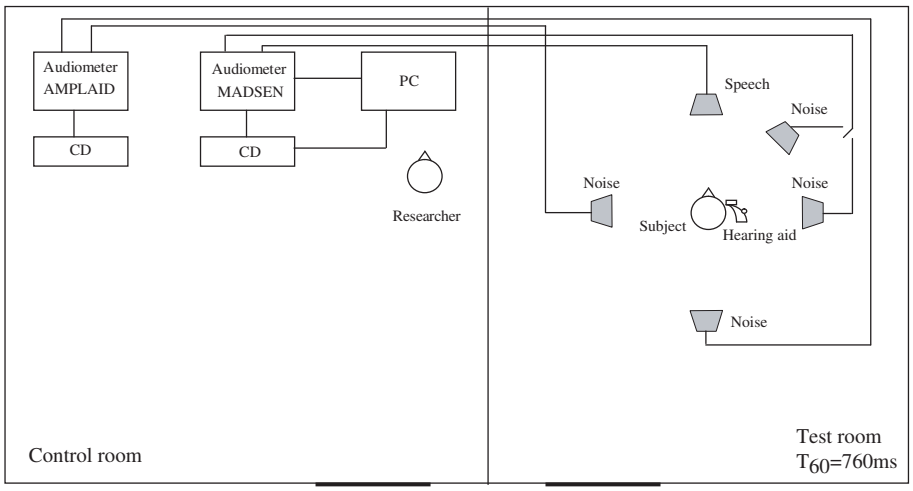


Figure 3.28: The experimental set-up for the perceptual evaluations of adaptive signal processing strategies implemented in the GNReSound Canta7 hearing aid.

through the MADSEN OB822 audiometer (channel left) and the noise at 180^o and 270^o through an AMPLAID 309 audiometer (channel left and right).

3.5.3 Physical evaluation

The physical evaluation involved acoustic measurements in the anechoic chamber with the hearing aid in stand-alone configuration or mounted on a mannequin, to calibrate the filters of the noise reduction algorithms and to calculate a directivity index (DI). Also acoustic measurements were performed in a reverberant chamber when the hearing aids are mounted on a mannequin, to calculate intelligibility-weighted polar diagrams.

3.5.3.1 Calibration

The noise reduction algorithms need to be calibrated for two different purposes, namely to match the level of the microphone signals at 1.6kHz and to calculate the coefficients of the fixed filters. The latter defines the spatial characteristics of the fixed directional microphone and the first filter of the adaptive beamformer. The calibration was carried out in the anechoic chamber with five different Canta7 hearing aids.

To match the level of both microphones of the hearing aid at 1.6kHz, a gain was applied to the rear omnidirectional microphone signal. To calculate this gain, the hearing aid was in stand-alone configuration and the loudspeaker (Yamaha CBX-S3) located at 90^o (relative to the endfire angle) at one meter in the azimuth plane of the hearing aid. The test signal was a one-third octave band noise with center frequency at 1.6kHz. An ANC algorithm with one coefficient updated by means of a NLMS procedure was used to determine the gain. The adaptation of the ANC was stopped after a few seconds and the obtained coefficient was used as the gain for the rear microphone. With the Audiologic software, the adaptation of the filter coefficients could be checked.

The fixed directional microphone had a hypercardioid polar diagram and the coefficients were defined to minimize the output signal when a source was present at an angle 110^o. For this, a test signal was presented by a loudspeaker located at an angle 110^o in the azimuth plane of the center of the mannequin. The test signal (ICRA-signal) was an unmodulated noise representative of a male weighted idealized speech spectrum at normal effort (ICRA, 1997). The filter coefficients were adapted by means of a NLMS procedure and the adaptation was stopped after a few seconds. The obtained coefficients were then used to create the fixed directional microphone. The coefficients of the first filter of the adaptive beamformer were defined to give a specific look direction to the algorithm, more specifically the coefficients were calculated to minimize the presence of the speech signal in the noise reference. The same procedure as for the fixed directional microphone was performed but now with the loudspeaker at 0^o.

3.5.3.2 Physical performance

In anechoic conditions, the evaluation was performed with the ICRA-signal. In reverberant conditions, a stationary speech-weighted noise was used, corresponding to Dutch sentences from a male speaker (also used in the perceptual evaluation). The signals of the omnidirectional microphone and the fixed directional microphone were directly accessible through the ear simulator set-up. The signal of the ADM and the A2B changed with the adaptation of the second filter. To carry out the physical evaluation of the adaptive algorithms, the coefficients were adapted on the noise signal until convergence and then they were kept fixed. Then, the reference signal (at 0^o) and noise signals were filtered separately with the fixed filters. These two separately filtered signals were then used for the physical evaluation of the adaptive directional microphone and the adaptive beamformer.

3.5.4 Perceptual evaluation

The tests with the three strategies (FDM, ADM and A2B) were carried out in four different noise scenarios (single noise source at 45^o, at 90^o, at 180^o and three uncorrelated noise sources at 90^o, 180^o and 270^o relative to speaker position). Two spectro-temporally different noise sounds (SW: unmodulated stationary speech weighted noise and MB: multitalker babble) and one speech material (sentences spoken by a male speaker) were used. Because of calibration issues all perceptual measurements were carried out with the same Canta7 hearing aid. In total, 608 SRT-measurements were carried out (480 tests and 128 test-retest).

3.5.4.1 Subjects

Fifteen normal hearing listeners participated in these measurements. Their ages ranged from nineteen to twenty three (mean of twenty one years) and their pure tone thresholds were less than or equal to 20dBHL at the octave frequencies from 125Hz to 8kHz. The perceptual evaluation was carried out in a monaural configuration. The monaural listening situation was created by an earplug (Bilsom 303S-30 or 303L-30) and by an earcap (Bilsom 2301) at the non-test ear.

3.5.4.2 Test materials

The adaptive procedure was used to extract the SRT in a constant level noise. The level step size for each speech material item was 2dB. In order to determine the SRT, the levels of the last ten responses of a list of thirteen sentences were averaged. The four noise sources were calibrated separately, to obtain a constant sound level at the reference point. Thus, a sound level of 65dBSPL and 69.8dBSPL was obtained with one noise source and three noise sources, respectively. Sentences spoken by a male speaker were used as speech materials. Unmodulated speech weighted noise (SW: with identical spectrum as the corresponding speech material) and multitalker babble (MB) were used as noise materials. The sentence speech materials were the Dutch sentences developed by Versfeld et al. (2000). One of the two test noises was unmodulated noise speech weighted according to the spectrum of the specific speech materials used. The other noise was multitalker babble, taken from the compact disk Auditory Tests (Revised) edited by Auditec of St. Louis.

3.5.5 Results and discussions

3.5.5.1 Physical evaluation

The calibration of the noise reduction algorithms was performed for five different hearing aids. To check the influence of the calibration on the noise

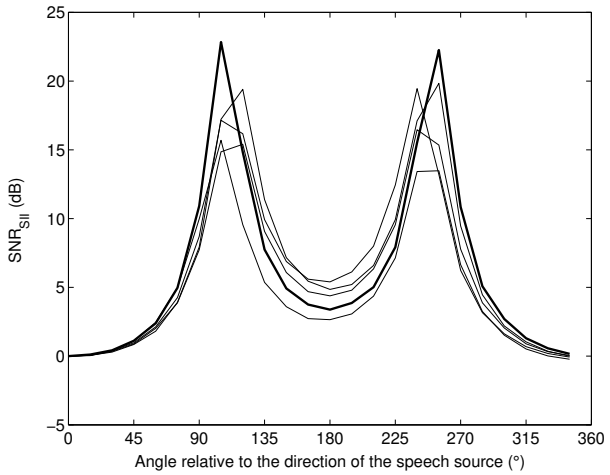


Figure 3.29: Intelligibility-weighted polar diagrams (in dB) of the fixed directional microphone for five different hearing aids. The calibration of one hearing aid was carried out (thick line) and this calibration was kept for the four other hearing aids. The plotted values are relative to the angle of the direction of the speech source (0^o), when the SNR is 0dB, measured on the basis of speech noise in stand-alone configuration and in anechoic conditions.

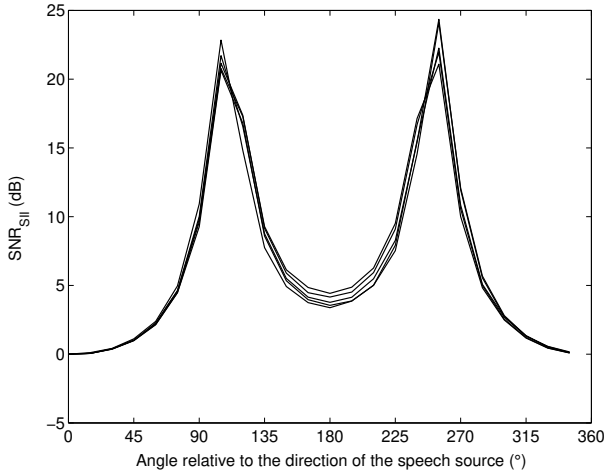


Figure 3.30: Intelligibility-weighted polar diagrams (in dB) of the fixed directional microphone for five different hearing aids. The calibration was carried out separately for the five different hearing aids. The plotted values are relative to the angle of the direction of the speech source (0^o), when the SNR is 0dB, measured on the basis of speech noise in stand-alone configuration and in anechoic conditions.

reduction performance, the intelligibility-weighted polar diagrams of the fixed directional microphone were measured in stand-alone and anechoic condition. In the first experiment, the calibration of the fixed directional microphone was performed with one hearing aid and this calibration was kept for the four other hearing aids. **Figure 3.29** shows the intelligibility-weighted polar diagrams of the five hearing aids and it can be seen that there were important variations in the shape of the polar diagrams. The non-smooth curves for the four hearing aids (slim line on **figure 3.29**) is due to the 15^o step size of the measurement procedure. The differences between the maximum and the minimum of the SNR_{SII} for the five hearing aids are 7.5dB, 2.9dB and 9.2dB at angles 105^o, 180^o and 255^o respectively. In the second experiment, the calibration of the fixed directional microphone was carried out for the five different hearing aids separately. The intelligibility-weighted polar diagrams of the five hearing aids are depicted in **figure 3.30**. The shapes of the polar diagrams are seen to be roughly similar. In this case, the differences between the maximum and the minimum of the SNR_{SII} for the five hearing aids are 1.9dB, 1.1dB and 3.3dB at angles 105^o, 180^o and 255^o respectively. This means that the differences in magnitude and phase between each microphone have an important impact on the noise reduction performance of the algorithm. For optimal use, a calibration of every algorithm has to be performed for every hearing aid. The performance of the noise reduction algorithm depends too much on the level of matching between the microphones used in the hearing aid. From now on, only one hearing aid is used for the physical and the perceptual evaluations.

The DI of each algorithm is presented in **figure 3.31**. As expected, the omnidirectional microphone (shown as a reference here) had the same sensitivity for all angles resulting in a DI approximately 0dB. At low frequencies (125-250Hz), the noise reduction schemes had low values for the DI. The ADM performed worse than the omnidirectional microphone below 170Hz. A negative DI means that the algorithm is more sensitive to the back hemisphere than to the front hemisphere. Above 170Hz, the A2B performed better than the FDM, which performed better than the ADM. Above 250Hz, the A2B performed at least 2dB better than the FDM. The difference between the FDM and the ADM strategy was only 1dB for frequencies above 450Hz. The DI_{AI} equaled 0.2dB for the omnidirectional microphone, 5.9dB for the FDM, 5.1dB for the ADM and 8.6dB for the A2B.

Figure 3.32 shows the intelligibility-weighted polar diagrams of the noise reduction algorithms and the front omnidirectional microphone in reverberant conditions. The interpretation of this polar diagram can be separated in three different hemispheres. The side of the hearing aid between 0^o and 105^o (ipsilateral), the back between 105^o and 255^o and the contralateral side between 255^o and 0^o. On the ipsilateral hemisphere, the noise reduction algorithms always performed better than the omnidirectional microphone. The FDM had roughly the same effect as the ADM. The A2B performed better

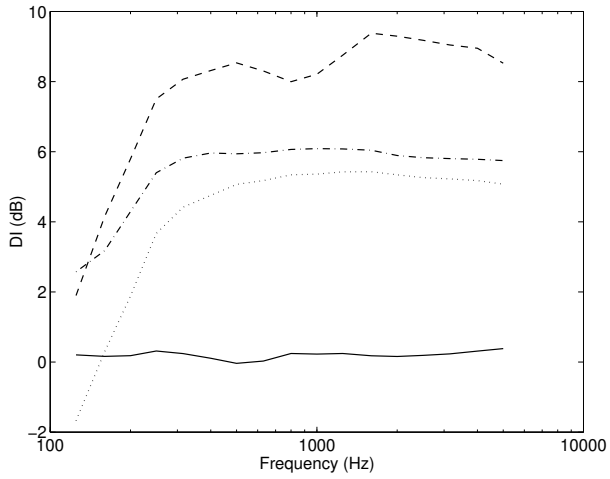


Figure 3.31: The Directivity Index DI (in dB) of the omnidirectional microphone (−), the fixed directional microphone (−·), the adaptive directional microphone (· · ·) and the two-stage adaptive beamformer (−−) are presented. The plotted values are relative to the angle of the direction of the speech source (0^o), when the SNR at the reference point is 0dB, measured on the basis of speech noise and the hearing aid stand-alone configuration and in anechoic conditions.

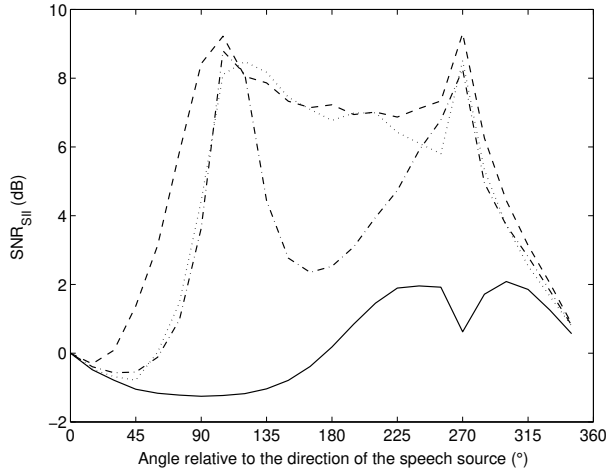


Figure 3.32: Intelligibility-weighted polar diagrams (in dB) of the omnidirectional microphone (−), the fixed directional microphone (−·), the adaptive directional microphone (· · ·) and the two-stage adaptive beamformer (−−). The angles are relative to the direction of the speech source (0^o). The SNR at the center of the head is 0dB, measured on the basis of speech noise, with the hearing aid on a mannequin and in reverberant conditions (T_{60}=0.76s).

than the other noise reduction algorithms especially between the angles 45^o and 90^o. This improvement was at least 2dB higher at 45^o up to 4dB at 90^o. In the back hemisphere, the ADM and the A2B technique had roughly the same performance except for angles from 210^o to 255^o where the A2B was slightly better than the ADM algorithm. The main difference in this hemisphere was between the adaptive algorithms and the FDM. At 180^o, the adaptive schemes performed about 4.4dB better than the fixed directional microphone. In the left hemisphere, the four algorithms had roughly the same performance in SNR. The A2B was always slightly better than the FDM and the ADM techniques but the improvement did not exceed 1dB.

3.5.5.2 Perceptual evaluation

Table 3.5 shows the improvements (in dB) of the SRT relative to the omnidirectional microphone for the fixed microphone (FDM) and the two adaptive techniques (ADM and A2B). The data correspond to the mean (and the standard deviation) of these improvements of all fifteen subjects for the four noise noise scenario configurations and the two noise materials.

An analysis of variance (ANOVA) was carried out using SPSS 10.0 software. The tests on within-subject analysis showed that there were significant differences between the noise materials ($F=508.8$, $p<0.001$), between the noise scenarios ($F=47.3$, $p<0.001$) and between the noise reduction algorithms ($F=49.1$, $p<0.001$). From tests of within-subject contrasts, the analysis showed that the interaction term between the algorithms and the noise materials was not significantly different from chance ($F=0.369$, $p=0.776$). This means that the noise reduction performance of the different techniques was independent of the noise materials. This allowed the averaging of the SRT improvements obtained with the stationary speech-weighted noise and multitalker babble noise tests. However, the absolute SNR at 50% intelligibility for the speech material (or SRT) was on average about 7.3dB higher for the multitalker babble than for the stationary speech weighted noise. Speech understanding in the multitalker babble noise was significantly more difficult than in stationary speech weighted noise. For each noise reduction technique, the improvements in SRT, relative to the omnidirectional microphone, depended on the noise scenarios. The performance of the noise reduction scheme was smallest when the noise source was close to the speaker, at 45^o. The best performance, relative to the omnidirectional microphone, was obtained for the FDM (4.8dB) and the A2B (8.2dB) with a noise source at 90^o, and for the ADM (5.3dB) with a noise source at 180^o. The four first subjects carried out all the SRT measurements two times on two different days. This was done to obtain an idea about the within-subject variability of the obtained results. The mean difference between both tests was 3.3dB with a standard deviation of 4.2dB. From a separate t-test analysis, it was shown that the first and the second measurement were not significantly different ($p=0.489$). To compare the performance of the noise re-

Table 3.5: The differences in SRT (dB) averaged for the fifteen normal hearing listeners (mean(SD)) of the fixed directional microphone (FDM), the adaptive directional microphone (ADM) and the adaptive beamformer (A2B), relative to the omnidirectional microphone, for different test conditions.

Noise configuration		FDM	ADM	A2B
45°	SW	0.8 (2.3)	0.2 (1.9)	2.2 (2.0)
	MB	1.9 (3.8)	-0.2 (2.8)	2.6 (2.4)
90°	SW	4.7 (3.3)	3.6 (2.8)	8.1 (3.2)
	MB	5.0 (3.9)	3.5 (3.5)	8.4 (2.9)
180°	SW	3.2 (3.0)	4.7 (3.3)	6.9 (2.7)
	MB	3.3 (3.8)	5.9 (4.1)	7.6 (3.8)
90°/180°/270°	SW	3.1 (2.8)	2.8 (2.8)	4.4 (2.5)
	MB	4.3 (3.3)	3.2 (3.4)	3.9 (2.7)

duction techniques, a paired comparison was carried out for the different noise configurations.

With a noise source at 45^o, there were no significant differences between the omnidirectional microphone, the FDM ($p=0.072$) and the ADM ($p=0.977$). Only the A2B performed significantly different from the omnidirectional microphone ($p<0.001$). However, there were no significant differences between the FDM and the A2B ($p=0.167$). Relatively to the omnidirectional microphone, an improvement in SRT of 1.3dB was found with the FDM and 2.4dB with the A2B. No improvement in SRT was obtained with the ADM. These results were in close agreement with the physical evaluation (**figure 3.32**).

With a noise source at 90^o, all the noise reduction techniques performed significantly different from the omnidirectional microphone ($p<0.001$). The largest improvement in SRT was found for the A2B, namely 8.2dB relatively to the omnidirectional microphone and 4.8dB with the FDM. These results agreed with results from a previous study (Maj et al., 2003) where similar improvements in SRT were obtained for the A2B (8.7dB) and for the FDM (4.7dB) for similar conditions (speech material, noise material and test room). With the ADM, an improvement of 3.5dB was obtained. There were no significant differences between the FDM and the ADM in a speech-weighted noise ($p=0.084$). This was predicted by the physical evaluation on **figure 3.32** at 90^o. The behaviour of the two techniques was roughly the same. However, in multitalker babble there were significant differences between these two noise reduction techniques ($p=0.018$). As already mentioned (Ricketts, 2000a; Maj et al., 2003), intelligibility-weighted polar diagrams as well as the directivity index have a strong link with the improvement of the speech intelligibility in noise. The FDM performed better than the ADM in this noise scenario because the noise source was close to 110^o, the angle where the FDM has a null.

The improvement in SRT (3.5dB) obtained by the ADM corresponded to the values measured for a software directional microphone in other studies (Maj et al., 2003). It seems that with the constraint on the adaptive filter, a noise source at 90^o did not fully allow the adaptation of the adaptive filter (section 3.5.1.2). The software directional microphone signal (which was connected to the speech reference of the algorithm and had a null at 180^o) was obtained at the output of the ADM.

For a single noise source at 180^o, all the strategies were different from each other ($p<0.001$). The A2B (7.2dB) gave the best improvement in speech intelligibility. An improvement of 3.2dB and 5.3dB was obtained with the FDM and the ADM, respectively. The ADM was better than the FDM approach in this noise scenario. The noise source was not close to the optimal angle (110^o) for the FDM and the noise source was at the back hemisphere, which means that the constraint allowed the adaptation in the ADM. In this way the adaptation brought an additional noise reduction to the software directional microphone of the ADM.

In a complicated noise scenario, with three uncorrelated noise sources, the noise reduction techniques all performed significantly different from the omnidirectional microphone ($p<0.001$). However, there were no significant differences between the noise reduction algorithms. Improvements in SRT of 3.7dB, 3.0dB and 4.1dB were obtained with the FDM, the ADM and the A2B, respectively. This means that speech understanding in noise is clearly enhanced by the different techniques but that the adaptive systems had the same effect as the FDM, i.e. the adaptive schemes did not bring a significant additional SRT-improvement in complex noise scenarios. This additionally stresses the approach, as confirmed by data (Greenberg and Zurek, 1992; Maj et al., 2003), that a good noise reduction scheme should have an adaptive processing for low reverberation or simple noise source scenarios and a fixed microphone directivity for high reverberation or complex noise scenarios. The FDM performed better than a conventional software directional microphone (Maj et al., 2003). On the one hand, the FDM has ten filter-taps and takes into account the characteristic of the microphones while on the other hand the software directional microphone (Thompson, 1999) has only one filter-tap and assumes that the characteristic of the microphones are matched. This can explain the difference in performance between both fixed directional microphones.

3.6 Conclusion

In this chapter, a two-stage adaptive beamformer was presented and further developed for hearing aid applications. The algorithm was optimized for implementation in digital hearing aids and the experiments demonstrated that

the proposed technique can outperform the noise reduction strategies actually implemented in current commercial hearing aids. The adaptive beamformer algorithm was assessed theoretically, physically and perceptually.

The *theoretical evaluation* showed that the parameters of the microphones such as the distance between the two sensors and the internal noise may have an important influence on the noise reduction performance of the A2B. The performance of the noise reduction algorithm improves by increasing the distance between the microphones and by decreasing the internal noise of the sensors. However, the space between the sensors is limited by the housing of the hearing aid and the internal noise is function of the microphones inside the device.

The *physical evaluation* demonstrated that the two-stage adaptive beamformer does not distort the speech signal when the speaker is in the look direction of the strategy (0^o). This was predicted by the theoretical evaluation. Important gains in intelligibility weighted SNR were measured for different directions of the jammer sound sources. Based on intelligibility weighted SNR and directivity measures, it was shown that the two-stage adaptive beamformer gives better noise reduction performance compared to software directional microphones, hardware directional microphones and adaptive directional microphones. From calibration issues, the experiments demonstrated that different hearing aids have different characteristics. For having a similar performance, individualized matching or calibration of the fixed parts of adaptive systems in two microphone hearing aids is necessary.

The *perceptual evaluations* revealed that a significant improvement of the speech intelligibility can be obtained with the two-stage adaptive beamformer. Although it was shown before that the effect of noise reduction strategies on speech intelligibility in noise could be very dependent on reverberation, these evaluation results were obtained in moderately reverberant environments and are thus representative of the effects in real living rooms. The improvement in SNR with the two-stage adaptive beamformer technique, as well as with other noise reduction approaches, of course also depends on the jammer sound scene. In simple jammer sound source scenarios the A2B always performs better than the omnidirectional microphone, the directional microphone and the adaptive directional microphone. However, in more complex noise scenarios, such as multiple jammer sound sources or diffuse noise, the performance of the adaptive beamformer falls back to the effect of the directional microphone. This is why a good noise reduction scheme should have an adaptive processing for low reverberation or simple noise scenarios and a robust fixed processing for high reverberation or complex noise scenarios. The perceptual results were in agreement with the physical evaluations indicating that very good prediction of SRT-improvement can be obtained using physical measurements in realistic conditions (on a mannequin, in realistic reverberation and SII calculation).

Chapter 4

SVD-based optimal filtering technique

In this chapter, an optimal filtering technique is presented and developed for dual microphone hearing aids. Although this strategy has a high computational complexity for implementation in digital hearing aids, the aim of this study is not to reduce the cost of calculation. This is the purpose of other studies. The objective of this chapter is to optimize and to evaluate the proposed method. The optimal filtering technique, which works without a priori assumptions about the signals of the microphones, is expected to give an additional gain in speech intelligibility relative to the two-stage adaptive beamformer in complex noise scenarios.

In the *first section* the optimal filtering technique is described and in the *second section* a theoretical evaluation is carried out. The influence of the characteristics of the microphones, the errors of the VAD and the jammer sound source scenarios on the noise reduction performance of the optimal filtering strategy and the two-stage adaptive beamformer is investigated. In the *third section*, a method to compute the optimal filtering strategy through a SVD-based technique is described. In the *fourth section*, a physical evaluation is performed. The parameters of the SVD-based optimal filtering method are optimized by maximizing the gain in intelligibility weighted SNR and experiments are carried out to assess the optimal filtering technique in different noise scenarios. The influence of the VAD errors on the performance of both noise reduction strategies is investigated. In the *fifth section*, a time constraint is introduced and a real time implementation of the SVD-based optimal filtering method is described. In the *sixth section* a method to improve the performance of the VAD is examined and assessed physically. Finally in a *seventh section*, a perceptual evaluation is carried out in a single and multiple noise scenario.

4.1 Optimal filtering technique

The noise reduction technique considered here, produces a speech signal estimate $\hat{\mathbf{s}}_k$ at time k from noisy data $\mathbf{u}_k = \mathbf{s}_k + \mathbf{n}_k$ by means of a linear filter $\mathbf{W}_k \in \mathbb{R}^{ML^{WF} \times ML^{WF}}$ (with L^{WF} the size of the filter per channel and M the number of microphones) using:

$$\hat{\mathbf{s}}_k = \mathbf{W}_k^T \mathbf{u}_k \tag{4.1}$$

where $\mathbf{u}_k \in \mathbb{R}^{ML^{WF}}$ is the filter input time series, and $\hat{\mathbf{s}}_k \in \mathbb{R}^{ML^{WF}}$ is the filter output time series at time k. The $\hat{\mathbf{s}}_k$ is then an estimate for the speech contribution $\mathbf{s}_k$ in $\mathbf{u}_k$ and the i-th column of $\mathbf{W}_k$ is a filter for the estimation of the i-th component of $\mathbf{s}_k$. In a two microphone application, the vector $\mathbf{u}_k$ is represented by:

$$\mathbf{u}_k = \begin{bmatrix} \mathbf{u}_{1k} & \mathbf{u}_{2k} \end{bmatrix}^T \tag{4.2}$$

with

$$\mathbf{u}_{jk} = \begin{bmatrix} u_{j,k} & u_{j,k-1} & \dots & u_{j,k-L^{WF}+1} \end{bmatrix}^T \tag{4.3}$$

where the j refers to the j-th microphone. The vectors $\mathbf{s}_k$ and $\mathbf{n}_k$ are similarly defined.

The linear filter $\mathbf{W}_k$ minimizes the estimation error between the clean speech $\mathbf{s}_k$ and the estimate of the clean speech $\hat{\mathbf{s}}_k$. To compute the filter $\mathbf{W}_k$ different cost functions are used such as the LS criterion (Hansen, 1997) and the MSE criterion (Doclo and Moonen, 2002). The latter is considered in this dissertation and the estimation error at time k is defined by $\mathbf{e}_k = \mathbf{s}_k - \hat{\mathbf{s}}_k$. The MSE cost function for obtaining an optimal filter is:

$$\begin{aligned} \mathbf{J}_{MSE} &= \mathcal{E}\{\|\mathbf{e}_k\|_2^2\} = \mathcal{E}\{\|(\mathbf{s}_k - \hat{\mathbf{s}}_k\|_2^2)\} \\ \mathbf{J}_{MSE} &= \mathcal{E}\{\mathbf{s}_k^H \mathbf{s}_k\} - 2\mathcal{E}\{\mathbf{u}_k^H \mathbf{W}_k \mathbf{s}_k\} + \mathcal{E}\{\mathbf{u}_k^H \mathbf{W}_k \mathbf{W}_k^H \mathbf{u}_k\} \end{aligned} \tag{4.4}$$

The optimal filter is found by setting the derivative $\frac{\partial \mathbf{J}_{MSE}}{\partial \mathbf{W}_k}$ equal to zero:

$$\frac{\partial \mathbf{J}_{MSE}}{\partial \mathbf{W}_k} = -2\mathcal{E}\{\mathbf{u}_k \mathbf{s}_k^H\} + 2\mathcal{E}\{\mathbf{u}_k \mathbf{u}_k^H\}\mathbf{W}_k = 0 \tag{4.5}$$

Hence, the optimal filter (or Wiener filter) $\mathbf{W}_{WF}$ is:

$$\mathbf{W}_{WF,k} = \mathcal{E}\{\mathbf{u}_k \mathbf{u}_k^H\}^{-1} \mathcal{E}\{\mathbf{u}_k \mathbf{s}_k^H\} \tag{4.6}$$

However, obtaining an estimate for $\mathcal{E}\{\mathbf{u}_k \mathbf{s}_k^H\}$ is not straightforward as $\mathbf{s}_k$, obviously, is unknown. If a robust VAD algorithm is used, and noise-only observations can be made during noise-only periods at time k' ($\mathbf{u}_{k'} = 0 + \mathbf{n}_{k'}$), then such observations can be used to estimate

$$\mathcal{E}\{\mathbf{n}_k \mathbf{n}_k^H\} = \mathcal{E}\{\mathbf{n}_{k'} \mathbf{n}_{k'}^H\} = \mathcal{E}\{\mathbf{u}_{k'} \mathbf{u}_{k'}^H\} \tag{4.7}$$

which means that $\mathcal{E}\{\mathbf{n}_k\mathbf{n}_k^H\}$ can be estimated. The leftmost equation indicates that short-term noise stationarity is assumed. During speech-and-noise periods (at time k), both the signal-of-interest and the noise signal are observed, $\mathcal{E}\{\mathbf{u}_k\mathbf{u}_k^H\}$. If $\mathbf{s}_k$ and $\mathbf{u}_k$ are assumed statistically independent of each other ($\mathcal{E}\{\mathbf{s}_k\mathbf{n}_k^H\} = 0$), then

$$\begin{aligned}\mathcal{E}\{\mathbf{u}_k\mathbf{u}_k^H\} &= \mathcal{E}\{\mathbf{s}_k\mathbf{s}_k^H\} + \mathcal{E}\{\mathbf{s}_k\mathbf{n}_k^H\} + \mathcal{E}\{\mathbf{n}_k\mathbf{s}_k^H\} + \mathcal{E}\{\mathbf{n}_k\mathbf{n}_k^H\} \\ &= \mathcal{E}\{\mathbf{n}_k\mathbf{n}_k^H\} + \mathcal{E}\{\mathbf{s}_k\mathbf{s}_k^H\}\end{aligned} \tag{4.8}$$

Given $\mathcal{E}\{\mathbf{u}_k\mathbf{u}_k^H\}$ and $\mathcal{E}\{\mathbf{n}_k\mathbf{n}_k^H\}$, $\mathcal{E}\{\mathbf{s}_k\mathbf{s}_k^H\}$ can be computed. Finally, from the assumed independence of $\mathbf{s}_k$ and $\mathbf{n}_k$ it also follows that

$$\mathcal{E}\{\mathbf{u}_k\mathbf{s}_k^H\} = \mathcal{E}\{\mathbf{s}_k\mathbf{s}_k^H\} + \mathcal{E}\{\mathbf{s}_k\mathbf{n}_k^H\} = \mathcal{E}\{\mathbf{s}_k\mathbf{s}_k^H\} \tag{4.9}$$

so that the optimal filter solution is finally given by

$$\begin{aligned}\mathbf{W}_{WF,k} &= \mathcal{E}\{\mathbf{u}_k\mathbf{u}_k^H\}^{-1}(\mathcal{E}\{\mathbf{u}_k\mathbf{u}_k^H\} - \mathcal{E}\{\mathbf{n}_{k'}\mathbf{n}_{k'}^H\}) \\ \mathbf{W}_{WF,k} &= \mathbf{R}_{u,u}^{-1}(\mathbf{R}_{u,u} - \mathbf{R}_{n,n})\end{aligned} \tag{4.10}$$

The computation of the optimal filter $\mathbf{W}_{WF,k}$ provides an estimator $\mathbf{w}_{WF,k}$ for an optimal filtered estimate $\tilde{s}_k$ at time k, but in addition optimal smoothed estimates $\tilde{s}_{k-1}, \tilde{s}_{k-2}, \ldots, \tilde{s}_{k-ML^{WF}+1}$ (all at time k). Conversely, optimal estimates for s_k are obtained at time k, $k+1$, $\ldots$, $k+ML^{WF}-1$, corresponding to different columns of $\mathbf{W}_{WF,k}$. The i-th column of $\mathbf{W}_{WF,k}$ can be considered as a two-channel filter $\mathbf{w}_{WF,k}$, where each microphone is filtered with an L^{WF}-taps filter $\mathbf{w}_{j,k}^{WF}$:

$$\mathbf{w}_{WF,k} = \begin{bmatrix}\mathbf{w}_{1,k}^{WF} \\ \mathbf{w}_{2,k}^{WF}\end{bmatrix} \tag{4.11}$$

Figure 4.1 shows the optimal filtering technique applied to a two microphone configuration in hearing aids. One type has two omnidirectional microphones OO_{Fo} and OO_{Ro} (**figure 4.1a**), which are used to create a software directional microphone OO_{Sd}. The delay τ and the weight a, of the weight factor for the back port $\delta_h(f) = a\, e^{-j2\pi f\tau}$, are chosen to give a hypercardioid spatial characteristic in anechoic conditions (null at 110^o). A second type has one hardware directional microphone DO_{Hd} and one omnidirectional microphone DO_{Fo} (**figure 4.1b**). In both cases, the output of the optimal filtering technique is the sum of the filtered version of the microphone signals.

The question then arises which column of $\mathbf{W}_{WF,k}$ gives the best estimate of $\mathbf{s}_k$. A *first solution* consists of computing the error covariance matrix. The estimation error $\mathbf{e}_k$ is defined as

$$\mathbf{e}_k = \mathbf{s}_k - \tilde{\mathbf{s}}_k = \mathbf{s}_k - \mathbf{W}_{WF,k}^T\mathbf{u}_k \tag{4.12}$$

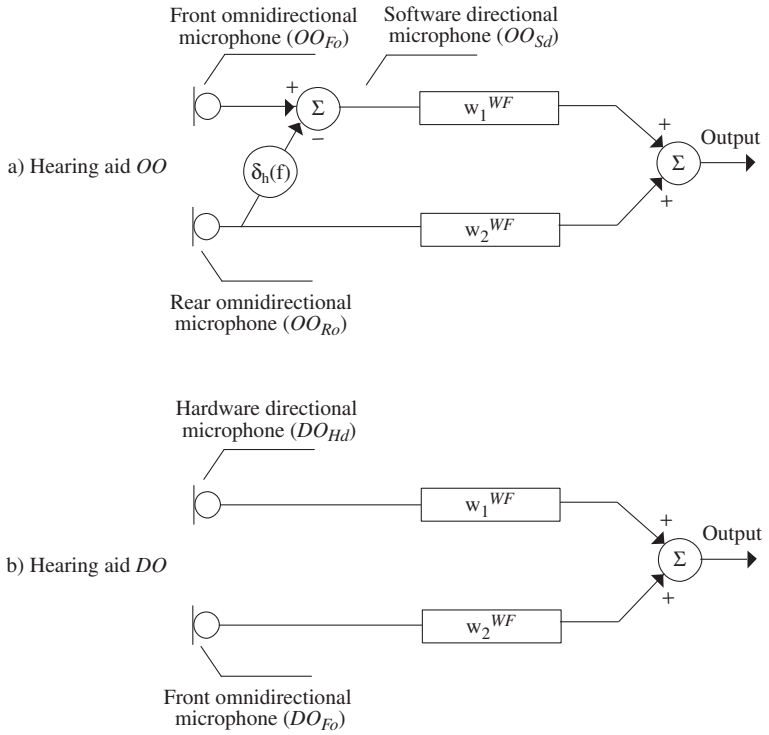


Figure 4.1: Representation of the optimal filter technique for the two types of hearing aids OO and DO. a) With hearing aid OO, two omnidirectional microphones (OO_{Fo} and OO_{Fo}), allows creating a software directional microphone OO_{Sd}. The rear omnidirectional microphone and the software directional microphone are connected to the optimal filtering technique. b) With hearing aid DO, one hardware directional microphone DO_{Hd} and one omnidirectional microphone DO_{Fo} are connected to the optimal filtering technique.

It is easily shown that the error covariance matrix (ECM) is given as

$$\mathcal{E}\{\mathbf{e}_k\mathbf{e}_k^H\} = \mathbf{R}_{n,n}\mathbf{W}_{WF,k} \tag{4.13}$$

In particular, the diagonal elements of the ECM $\{\mathbf{R}_{n,n}\ \mathbf{W}_{WF,k}\}^{i,i}$ are interesting. Indeed, the smallest element on the main diagonal of this error covariance matrix corresponds to the best estimator $\mathbf{w}_1 \triangleq \mathbf{W}_{WF,k}(:,i')$, where $i' = \arg\min_i(\mathcal{E}\{\mathbf{e}_k\mathbf{e}_k^H\})^{i,i}$. A *second solution* consists of computing an average over all available estimates. This technique is often applied to rank truncation based estimation (Hansen, 1997; Hansen and Jensen, 1998). The corresponding

filter for a channel is then given as

$$\mathbf{w}_2 \triangleq \left[\frac{1}{L^{WF}}\ \frac{1}{L^{WF}} \cdots \frac{1}{L^{WF}}\ \frac{1}{L^{WF}}\right] \cdot \tag{4.14}$$
$$\begin{bmatrix} 0 & 0 & \cdots & 0 & w(1,1) & \cdots & w(L^{WF}-1,1) & w(L^{WF},1) \\ 0 & 0 & \cdots & w(1,2) & w(2,2) & \cdots & w(L^{WF},2) & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \\ 0 & w(1,L^{WF}-1) & \cdots & w(L^{WF}-2,L^{WF}-1) & w(L^{WF}-1,L^{WF}-1) & \cdots & 0 & 0 \\ w(1,L^{WF}) & w(2,L^{WF}) & \cdots & w(L^{WF}-1,L^{WF}) & w(L^{WF},L^{WF}) & \cdots & 0 & 0 \end{bmatrix}^T$$

With $w(i,j)$ denoting the (i,j)-element of $\mathbf{W}_{WF,k}$. A *third solution* is to simply take an arbitrary column of $\mathbf{W}_{WF,k}$ and use this as a noise reduction filter, $\mathbf{w}_3 \triangleq \mathbf{W}_{WF,k}(:,i)$ for random $i \in \{1,\ldots,ML^{WF}\}$. A comparison between these three different strategies will be given in a section 4.4.1.3.1.

4.2 Theoretical evaluation

In this section, the theoretical evaluation of the optimal filtering technique is carried out and compared to the theoretical evaluation of the A2B (chapter 3). The equations of the transfer function between the input and the output of the optimal filtering technique are derived as a function of the CCF. Both noise reduction strategies are applied to the hearing aid OO (**figure 4.1a** and **3.4a**), and for brevity, the microphone indices OO are omitted in the equations. The equations of the A2B are derived in the section 3.2. The noise reduction performances of both noise reduction techniques are then evaluated by varying different parameters such as the characteristics of the microphones, the errors of the VAD and the jammer sound source scenarios. The publication related to this section is (Maj et al., 2002b).

4.2.1 Derivation of the equations

In the section 3.2, the CCF for a speech source at the endfire angle 0^o and with one noise source at broadside angle 90^o was given as:

$$\Gamma^{speech}_{Fo,Ro}(f) = exp(j\ 2\pi\ f\ d/c) \quad \Gamma^{noise}_{Fo,Ro}(f) = \frac{1}{1+\rho(f)} \tag{4.15}$$

and the PSD and the C-PSD as a function of the CCF at the output of the software directional microphone (OO_{Sd}) were given as:

$$\begin{aligned} P_{Sd,Sd}(f) &= P_{in,in}(f)\ (1 - 2\ Re(\delta^*_h(f)\ \Gamma_{Fo,Ro}(f)) + \delta_h(f)\ \delta^*_h(f)) \\ P_{Sd,Ro}(f) &= P_{in,in}(f)\ (\Gamma_{Fo,Ro}(f) - \delta_h(f)) \end{aligned} \tag{4.16}$$

when the PSDs of the received signals at the microphone input were assumed to be the same ($P^{noise}_{Ro,Ro}(f) = P^{noise}_{Fo,Fo}(f) \triangleq P^{noise}_{In,In}(f)$ and $P^{speech}_{Ro,Ro}(f) = P^{speech}_{Fo,Fo}(f) \triangleq P^{speech}_{In,In}(f)$).

With the hearing aid OO, the microphone inputs of the optimal filtering strategy (**figure 4.1a**) are the software directional microphone (OO_{Sd}) and the rear omnidirectional microphone (OO_{Ro}). The vectors containing successive time samples of microphone signals are $\mathbf{u}_k = [\mathbf{u}_{Sd,k} \quad \mathbf{u}_{Ro,k}]^T$ and $\mathbf{n}_k = [\mathbf{n}_{Sd,k} \quad \mathbf{n}_{Ro,k}]^T$. In the frequency domain, the correlation matrices $\mathbf{R}_{u,u}$ and $\mathbf{R}_{n,n}$ (equation (4.10)) correspond to:

$$\mathbf{R}_{u,u}(f) = \begin{bmatrix} P^{speech}_{Sd,Sd}(f) + P^{noise}_{Sd,Sd}(f) & P^{speech}_{Sd,Ro}(f) + P^{noise}_{Sd,Ro}(f) \\ P^{speech*}_{Sd,Ro}(f) + P^{noise*}_{Sd,Ro}(f) & P^{speech}_{Ro,Ro}(f) + P^{noise}_{Ro,Ro}(f) \end{bmatrix} \tag{4.17}$$

$$\mathbf{R}_{n,n}(f) = \begin{bmatrix} P^{noise}_{Sd,Sd}(f) & P^{noise}_{Sd,Ro}(f) \\ P^{noise*}_{Sd,Ro}(f) & P^{noise}_{Ro,Ro}(f) \end{bmatrix} \tag{4.18}$$

The computation of the optimal filter $\mathbf{W}_{WF,k}$ provides estimators $\mathbf{w}_{WF,k}$ for the different signals $\hat{\mathbf{s}}_k$. Section 4.4.2.1.1 shows that using the directional part of $\mathbf{W}_{WF,k}$ gives a good estimate of $\mathbf{s}_k$ because the SNR at the directional microphone is higher than the SNR at the omnidirectional microphone,

$$\hat{\mathbf{s}}_k = [\mathbf{w}^{WF}_{1,k} \quad \mathbf{w}^{WF}_{2,k}] \begin{bmatrix} \mathbf{u}_{Sd,k} \\ \mathbf{u}_{Ro,k} \end{bmatrix} \tag{4.19}$$

Thus, the two-channel estimator $\mathbf{w}_{WF}(f) = [\mathbf{w}^{WF}_1(f) \quad \mathbf{w}^{WF}_2(f)]^T$ is given as:

$$\begin{aligned} \mathbf{w}^{WF}_1(f) = \frac{1}{Det} \; & [P^{speech}_{Sd,Sd}(f)(P^{speech}_{Ro,Ro}(f) + P^{noise}_{Ro,Ro}(f)) \\ & - P^{speech}_{Sd,Ro}(f)(P^{speech*}_{Sd,Ro}(f) + P^{noise*}_{Sd,Ro}(f))] \end{aligned} \tag{4.20}$$

$$\begin{aligned} \mathbf{w}^{WF}_2(f) = \frac{1}{Det} \; & [P^{speech}_{Sd,Ro}(f)(P^{speech}_{Sd,Sd}(f) + P^{noise}_{Sd,Sd}(f)) \\ & - P^{speech}_{Sd,Sd}(f)(P^{speech}_{Sd,Ro}(f) + P^{noise}_{Sd,Ro}(f))] \end{aligned} \tag{4.21}$$

where Det corresponds to the determinant of the matrix $\mathbf{R}_{u,u}(f)$. The PSD at the output of the optimal filtering technique corresponds to:

$$\begin{aligned} P_{Out,Out}(f) = & |\mathbf{w}^{WF}_1(f)|^2 P_{Sd,Sd}(f) + |\mathbf{w}^{WF}_2(f)|^2 P_{Ro,Ro}(f) \\ & + 2 \; Re(\mathbf{w}^{WF}_1(f) P_{Sd,Ro}(f) \; \mathbf{w}^{WF*}_2(f)) \end{aligned} \tag{4.22}$$

4.2.2 Experimental set-up

As for the two-stage adaptive beamformer, the performance of optimal filtering strategy is a function of the sensor-to-environmental noise ratio $\rho(f)$,

the distance between the two microphones d and the signal-to-noise ratio $\alpha(f)$ (equation (4.15)). The characteristics of the microphones and the VAD errors may also influence the performance of both noise reduction strategies. A deviation in gain ($\Delta gain$) and/or in phase ($\Delta phase$) generally occurs between the microphones of the hearing aids. In the theoretical evaluation, the deviation in gain and/or phase, is applied on the rear omnidirectional microphone (OO_{Ro}) such that:

$$OO_{Ro}^{Dev}(f) = OO_{Ro}(f)\ \Delta gain\ e^{(j\Delta phase\ \pi/180)} \tag{4.23}$$

The errors of the VAD can be classified into two different classes. In the first class (equation (4.24)), the noise-only periods are considered as speech-and-noise periods, and in the second class (equation (4.25)), the speech-and-noise periods are considered as noise-only periods. Thus, the statistics of the speech (or the noise) signal are corrupted by the statistics of the noise (or the speech) signal. For modeling the errors of the VAD, the PSDs of the received signals at the microphone input for speech and noise ($P_{in,in}^{speech}$ and $P_{in,in}^{noise}$) are redefined as:

$$P_{VADerror}^{speech}(f) = (1 - VAD_{error})P_{in,in}^{speech}(f) + VAD_{error}P_{in,in}^{noise}(f) \tag{4.24}$$

$$P_{VADerror}^{noise}(f) = (1 - VAD_{error})P_{in,in}^{noise}(f) + VAD_{error}P_{in,in}^{speech}(f) \tag{4.25}$$

where VAD_{error} denotes the degree of the corruption. When $VAD_{error} = 0$, there is a perfect voice activity detection. With the two-stage adaptive beamformer, only the equation (4.25) is considered (a speech-and-noise period is detected as a noise-only period) because if a noise-only period is detected as a speech-and-noise period the coefficients of the A2B are kept fixed. This should affect the performance of the adaptive beamformer only in the case where the coefficients are not updated for a long period of time just before a speech-and-noise period, especially in non-stationary conditions. In this case, the statistics of the noise signal are not well evaluated and then a suboptimal estimation of the noise is subtracted from the speech and noise signal by the noise reduction algorithm.

The influence of multiple noise sources on the noise reduction strategies can be also assessed. The CCF function is the average of the CCF calculated at different angles. The noise sources were located at the angles 90^o, $90^o/180^o$ and $90^o/180^o/270^o$ for one, two and three noise source scenarios respectively.

4.2.3 Results

Simulations were carried out by varying the different parameters such as the distance d between the two microphones, the sensor-to-environmental noise ratio $\rho(f)$ and the signal-to-noise ratio $\alpha(f)$. These experiments were performed under the assumption that there was no deviation between the two microphones ($\Delta gain = 1$ and $\Delta phase = 0$) and no VAD errors ($VAD_{error} = 0$).

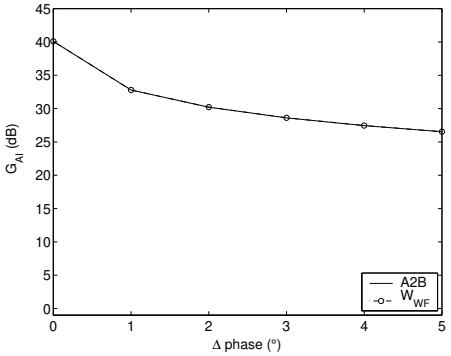


Figure 4.2: Influence of the phase deviation ($\Delta phase$) of the microphones on the G_{AI} for the two-stage adaptive beamformer and the optimal filtering technique ($d = 2cm$, $\alpha = 0$dB, $\rho = -45$dB).

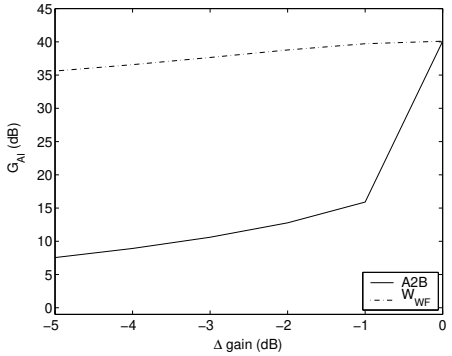


Figure 4.3: Influence of the gain deviation ($\Delta gain$) of the microphones on the G_{AI} for the two-stage adaptive beamformer and the optimal filtering technique ($d = 2cm$, $\alpha = 0$dB, $\rho = -45$dB).

A similar behaviour as for the two-stage adaptive beamformer (section 3.2.2) was found with the optimal filtering technique. The results are omitted in order to be concise.

Figures 4.2 and **4.3** illustrate the behaviour of the noise reduction algorithms (A2B and optimal filtering) with deviations in phase and gain ($d = 2cm$, $\rho = -45dB$ and $\alpha = 0dB$). With phase deviations, the G_{AI} decreased for both noise reduction algorithms. With gain deviations, on the one hand the performance of the G_{AI} of the A2B decreased significantly, while on the other hand, the SVD-based technique seemed to be robust against gain deviations.

Figures 4.4 and **4.5** show the effect of the VAD errors for both algorithms with no deviation in phase and gain ($d = 2cm$, $\rho = -45dB$ and $\alpha = 0dB$). These experiments were carried out for four different positions of the speaker. With the two-stage adaptive beamformer, the G_{AI} performance depended on the speaker position. Indeed, when the speaker was at 0^o, the performance of the beamformer technique was not affected. However, when the speaker was not at 0^o, the performance of the beamformer also decreased as a function of the VAD errors but not as drastically as for the optimal filtering approach. If the speaker is not positioned at the look direction (angle 0^o) of the beamformer, a leakage of the speech signal into the noise reference is obtained. **Figure 4.6** shows the level of the speech signal in the noise reference. The farther the speaker is away from position 0^o, the higher the level of the speech in the noise reference. In a first experiment with the optimal filtering technique, the noise statistic was corrupted by the speech statistic (equation

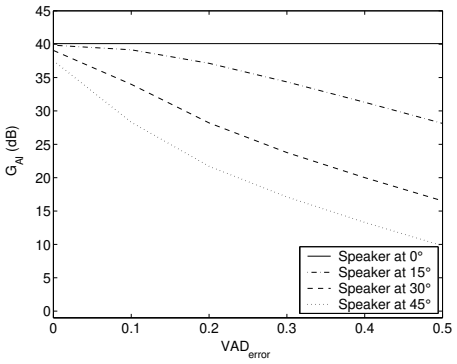


Figure 4.4: Influence of VAD errors on the G_{AI} for the two-stage adaptive beamformer ($\Delta gain = 0$dB, $\Delta phase = 0^o$, $d = 2cm$, $\alpha = 0$dB, $\rho = -45$dB).

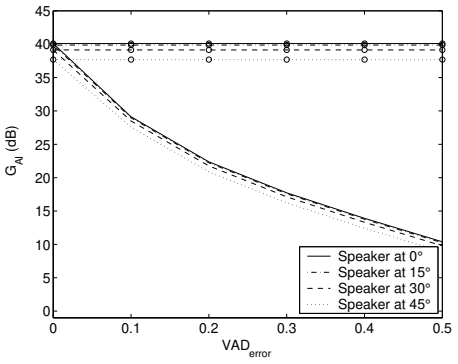


Figure 4.5: Influence of VAD errors on the G_{AI} for the optimal filtering technique when the noise statistics are corrupted by the speech statistics ($-$) and when the speech statistics are corrupted by the noise statistics (o) ($\Delta gain = 0$dB, $\Delta phase = 0^o$, $d = 2cm$, $\alpha = 0dB$, $\rho = -45$dB).

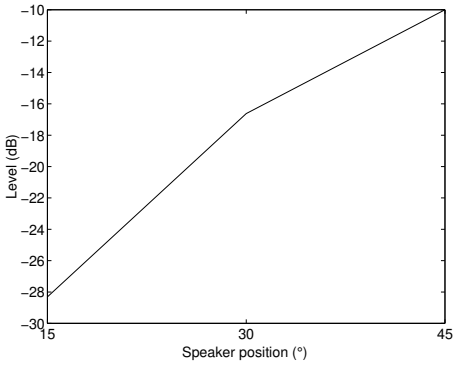


Figure 4.6: Level (dB) of the speech signal in the noise reference of the two-stage adaptive beamformer as a function of the speaker position ($\Delta gain = 0$dB, $\Delta phase = 0^o$, $d = 2cm$, $\alpha = 0$dB, $\rho = -45$dB).

(4.25)). The G_{AI} performance of the optimal filtering technique was independent of the speaker position but dropped rapidly as a function of the VAD errors **figure 4.5**. In a second experiment with the optimal filtering technique, the speech statistic was corrupted by the noise statistic (equation (4.24)). The performance of the technique was not dependent on the speaker position and

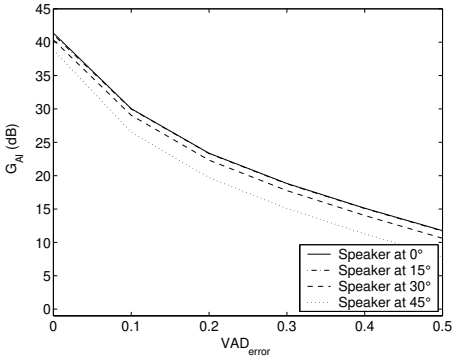


Figure 4.7: Influence of VAD errors (combined with a gain deviation) on the G_{AI} for the two-stage adaptive beamformer ($\Delta gain = -2.5$dB, $\Delta phase = 0^o$, $d = 2cm$, $\alpha = 0$dB, $\rho = -45$dB).

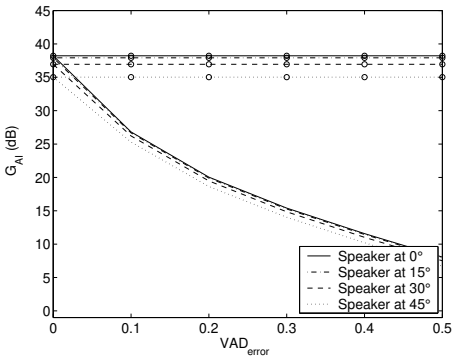


Figure 4.8: Influence of VAD errors (combined with a gain deviation) on the G_{AI} for the optimal filtering technique when the noise statistics are corrupted by the speech statistics (−) and when the speech statistics are corrupted by the noise statistics (o) ($\Delta gain = -2.5$dB, $\Delta phase = 0^o$, $d = 2cm$, $\alpha = 0$dB, $\rho = -45$dB).

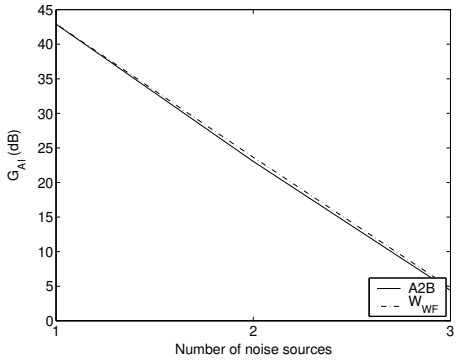


Figure 4.9: Influence of the multiple noise sources on G_{AI} for the two-stage adaptive beamformer and the optimal filtering technique ($\Delta gain = -2.5$dB, $\Delta phase = 0^o$, $d = 2cm$, $\alpha = 0$dB, $\rho = -45$dB).

not affected by the VAD errors (**figure 4.5**).

Another experiment combined a gain deviation ($gain = -2.5dB$) of OO_{Ro} with VAD errors (**figures 4.7** and **4.8**). The optimal filtering tech-

nique kept the same performance as in the last experiment, unlike the adaptive beamformer technique. Indeed, the performance dropped significantly, and the A2B had roughly the same behaviour as the optimal filtering technique. Furthermore, the farther the position of the speaker from 0^o, the worse the performance of the A2B.

Figure 4.9 shows the influence of multiple noise sources on the noise reduction performance of the strategies. Both algorithms had the same behaviour and the higher the number of jammer sound sources, the worse the noise reduction performances. Theoretically, an adaptive signal processing strategy comprised of N microphones can separate up to N statistically independent sources. Therefore, in the case of one desired source (the speaker) only one noise source can be cancelled (with appropriate processing) in a two-microphone set-up.

4.3 SVD-based optimal filtering technique

The computation of the optimal filtering technique through a GSVD-based technique is presented in this section. Doclo and Moonen (2002) used an interesting and convenient method to compute the optimal filter $\mathbf{W}_{WF}$ presented in formula (4.10). $\mathbf{W}_{WF}$ is derived from the joint diagonalization (*generalized eigenvalue decomposition*[1]) of the symmetric matrices $\mathbf{R}_{u,u} = \mathcal{E}\{\mathbf{u}_k \mathbf{u}_k^H\}$ and $\mathbf{R}_{n,n} = \mathcal{E}\{\mathbf{n}_{k'} \mathbf{n}_{k'}^H\}$,

$$\begin{cases} \mathbf{R}_{u,u} = \bar{\mathbf{X}}\ diag\{\bar{\sigma}_i^2\}\ \bar{\mathbf{X}}^T \\ \mathbf{R}_{n,n} = \bar{\mathbf{X}}\ diag\{\bar{\eta}_i^2\}\ \bar{\mathbf{X}}^T \end{cases} \tag{4.26}$$

where $\bar{\mathbf{X}}$ is a matrix containing the generalized eigenvalue vectors of $\mathbf{R}_{u,u}$ and $\mathbf{R}_{n,n}$. This matrix is an invertible matrix but not necessarily orthogonal. $\bar{\sigma}$, $\bar{\eta}$ are the generalized eigenvalues of $\mathbf{R}_{u,u}$ and $\mathbf{R}_{n,n}$. Note that $diag\{\bar{\sigma}_i^2\}$ represents a diagonal matrix with diagonal elements $\bar{\sigma}_i^2$, i=1 $\ldots ML^{WF}$, and that $diag\{\bar{\eta}_i^2\}$ is similarly defined.

In practice, $\bar{\mathbf{X}}$, $\bar{\sigma}_i^2$, $\bar{\eta}_i^2$ are approximated by means of a generalized singular value decomposition of the data matrices $\mathbf{U}_k \in \mathbb{R}^{p \times ML^{WF}}$ and $\mathbf{N}_k \in \mathbb{R}^{q \times ML^{WF}}$ (with p and q typically larger than ML^{WF}), such that

$$\mathbf{R}_{u,u} \Rightarrow (\mathbf{U}_k^T \mathbf{U}_k)/p \quad \mathbf{R}_{n,n} \Rightarrow (\mathbf{N}_k^T\ \mathbf{N}_k)/q \tag{4.27}$$

meaning that $\mathbf{R}_{u,u}$ and $\mathbf{R}_{n,n}$ are approximated because p and q are finite. The

[1] In appendix B, some definitions of linear algebra and matrix decompositions are given.

expanded form of the data matrices $\mathbf{U}_k$ and $\mathbf{N}_k$ are:

$$\mathbf{U}_k = \begin{bmatrix} \mathbf{u}_k^T \\ \mathbf{u}_{k+1}^T \\ \vdots \\ \mathbf{u}_{k+p-1}^T \end{bmatrix} \qquad \mathbf{N}_k = \begin{bmatrix} \mathbf{n}_{k'}^T \\ \mathbf{n}_{k'+1}^T \\ \vdots \\ \mathbf{n}_{k'+q-1}^T \end{bmatrix} \tag{4.28}$$

The data in $\mathbf{u}_k$ ($\mathbf{n}_k$) are collected during speech-and-noise (noise-only) periods. The generalized singular value decomposition of the matrices $\mathbf{U}_k$ and $\mathbf{N}_k$ is defined as:

$$\begin{cases} \mathbf{U}_k = \mathbf{Y}\ diag\{\sigma_i\}\ \mathbf{X}^T \\ \mathbf{N}_k = \mathbf{Z}\ diag\{\eta_i\}\ \mathbf{X}^T \end{cases} \tag{4.29}$$

where $\mathbf{Y} \in \mathbb{R}^{p \times ML^{WF}}$ and $\mathbf{Z} \in \mathbb{R}^{q \times ML^{WF}}$ are orthogonal matrices, $\mathbf{X} \in \mathbb{R}^{ML^{WF} \times ML^{WF}}$ is an invertible matrix and σ_i and η_i are the generalized singular values. By substituting the above formulas in (4.10), an alternative expression for $\mathbf{W}_{WF,k}$ is obtained:

$$\mathbf{W}_{WF,k} = \mathbf{X}^{-T}\ diag\left\{ \frac{q\ \sigma_i^2 - p\ \eta_i^2}{q\ \sigma_i^2} \right\} \mathbf{X}^T \tag{4.30}$$

In a two microphone case, the data matrices $\mathbf{U}_k \in \mathbb{R}^{(p \times 2L^{WF})}$ and $\mathbf{N}_k \in \mathbb{R}^{(q \times 2L^{WF})}$ as defined in equation (4.28) are block Toeplitz and take the form

$$\mathbf{U}_k = \begin{bmatrix} \mathbf{U}_{1k} & \mathbf{U}_{2k} \end{bmatrix} \qquad \mathbf{N}_k = \begin{bmatrix} \mathbf{N}_{1k} & \mathbf{N}_{2k} \end{bmatrix} \tag{4.31}$$

where

$$\mathbf{U}_{jk} = \begin{bmatrix} u_{j,k} & u_{j,k-1} & u_{j,k-2} & \cdots & u_{j,k-L^{WF}+1} \\ u_{j,k+1} & u_{j,k} & u_{j,k-1} & \cdots & u_{j,k-L^{WF}+2} \\ \vdots & \vdots & \vdots & & \vdots \\ u_{j,k+p-1} & u_{j,k+p-2} & u_{j,k+p-3} & \cdots & u_{j,k+p-L^{WF}} \end{bmatrix} \tag{4.32}$$

$$\mathbf{N}_{jk} = \begin{bmatrix} n_{j,k'} & n_{j,k'-1} & n_{j,k'-2} & \cdots & n_{j,k'-L^{WF}+1} \\ n_{j,k'+1} & n_{j,k'} & n_{j,k'-1} & \cdots & n_{j,k'-L^{WF}+2} \\ \vdots & \vdots & \vdots & & \vdots \\ n_{j,k'+p-1} & n_{j,k'+p-2} & n_{j,k'+p-3} & \cdots & n_{j,k'+p-L^{WF}} \end{bmatrix} \tag{4.33}$$

4.4 Physical evaluation

In the present section, the physical evaluation of the SVD-based optimal filtering technique is performed. In the first subsection, the experimental

set-up is presented. The hearing aid, the signals and the methods for optimizing the different parameters of the optimal filtering technique are described. The parameters of the optimal filtering technique are the particular choice of the optimal filter estimate ($\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$ in section 4.1), the length of the data matrices $\mathbf{U}_k$ and $\mathbf{N}_k$ (p and q) and the size L^{WF} per channel of the filter $\mathbf{w}_{WF,k}$. The set-up to assess the performance of the noise reduction strategy, for different noise source scenarios and VAD errors, is examined. In the second subsection, the optimal filtering technique is compared to the two-stage adaptive beamformer in different noise scenarios and the results are presented. The publication related to this section is (Maj et al., 2002a).

4.4.1 Experimental set-up

4.4.1.1 Hearing aid

The hearing aid described in section 3.1 referred as OO was used in these experiments. This hearing aid had a Danavox-163D housing where two omnidirectional microphones were mounted in an endfire array configuration spaced two centimeters apart (**figure 3.2**). The device had a linear amplifier and did not have systems for compression or feedback control. The block schemes of the optimal filtering technique and the two-stage adaptive beamformer for the hearing aid OO are illustrated in **figure 3.4a** and **4.1a**, respectively.

4.4.1.2 Signals

The two neighboring rooms described in section 3.4.1 were used for the experiments. In the test room, the hearing aid was positioned on a dummy head and loudspeakers (Yamaha CBX-S3) were situated at one meter from the mannequin.

The signals of the two microphones of the hearing aid were amplified and connected to the second room where further signal processing took place. The signal of the microphones were amplified with a two-channel Larson Davis 2200C amplifier and were digitized, at a sampling frequency of 16kHz, with a PC-platform using a Texas Instruments TMS320C40 digital signal processor (DSP) and two input channels with 16-bit analog-to-digital conversion (ADC). The amplifier of the hearing aid and the different devices were checked to ensure that they were not saturated.

The speech and noise signals, which were sent to the loudspeakers, were stationary speech weighted noises, namely ICRA, BLU, NVA and BRUGSE. The ICRA-signal was an unmodulated random gaussian noise representative of a male weighted idealized speech spectrum with normal effort (ICRA, 1997). The BLU-signal, NVA-signal and BRUGSE-signal are described and available on the Compact-Disk of Wouters (Wouters et al., 1994). They are also sta-

tionary speech-weighted noises but they do not have the same spectrum as the ICRA signal.

4.4.1.3 Optimization of the parameters

In these experiments, one loudspeaker was situated in front of the dummy head (at 0^o) and presented the speech material, and a second at 90^o (on the side of the hearing aid) presented the noise material.

The ICRA signal was used at 0^o for speech and a similar ICRA signal at 90^o for noise, implying that the noise and speech signals had the same spectrum. These two signals were recorded separately, i.e. the speech was recorded when only the loudspeaker in front of the dummy head was active, and the noise was recorded when only the loudspeaker on the side of the dummy head was active. This provided additional experimental flexibility. The two loudspeakers were matched to get a level of 70dBSPL at the center of the dummy head. The signals were recorded during ten seconds, thus two signals gave 160000 samples for speech as well as for noise. From these signals, a speech-and-noise frame (F_{SN}) by the addition of the speech and the noise signals, and a noise-only frame (F_N) was created only with the noise signal.

A point was defined inside the original 160000 samples frames referred to as *'start speech'*. *'Start speech'* could be interpreted as a perfect VAD and meant that before the point *'start speech'* there was a noise-only period (the noise-only frame was used) and after this point, there was a speech-and-noise period (the speech-and-noise frame was used).

With the SVD-based optimal filtering algorithm, the optimal filter $\mathbf{w}_{WF,k}$ is a two-channel filter, where each microphone is filtered with a L^{WF}-taps filter $\mathbf{w}_{j,k}^{WF}$, see formula (4.11). The resulting estimated signal is computed by summing the microphone signals $\mathbf{s}_k$ and $\mathbf{n}_k$ and filtering with the filters $\mathbf{w}_{j,k}^{WF}$ (**figure 4.1**). The G_{SII} computation used these two filtered signals separately to make the evaluation of the algorithm performance. For the two-stage adaptive beamformer the coefficients of the second filter were adapted on the noise signal until convergence and then they were kept fixed. Then, the speech and noise signals were filtered separately with the fixed coefficients, and the evaluation of the G_{SII} performance was measured in the same fashion as for the optimal filtering technique.

4.4.1.3.1 Optimal estimate

As already mentioned, $\mathbf{W}_{WF,k}$ (formula (4.10)) contains a collection of $2L^{WF} - tap$ filters. A choice for $\mathbf{w}_{WF,k}$ need to be made to obtain an estimate for $\mathbf{s}_k$. In paragraph 4.1, three different methods were described, an error-covariance matrix based method (ECM), an averaging over all available estimates and an arbitrary choice between the columns of the filter matrix

$\mathbf{W}_{WF,k}$. The SVD-based optimal filtering algorithm was applied with a speech-and-noise frame and noise-only frame of 2000 samples and the filter length L^{WF} was twenty per channel.

4.4.1.3.2 Influence of the design parameters

The optimal filtering technique depends on three parameters: the size of the noise-only frame (q), the size of the speech-and-noise frame (p), and the length of the filter for each channel (L^{WF}). In the next paragraph, the experiments are described to study the influence of these parameters on the speech intelligibility performance through the G_{SII} .

The signals $\mathbf{u}_k$ and $\mathbf{n}_k$ were stationary, which implied that for different values of *'start speech'*, the same G_{SII} improvement could be expected. The optimal filtering was applied for fourteen different values of *'start speech'*. With the beamformer, the second stage adapted on the 10000 samples of the noise-only frame before the *'start speech'*. The standard deviation and the mean of the improvement were calculated with the fourteen different values for *'start speech'* for both algorithms. Also, with the SVD-based optimal filtering technique, the length of the speech-and-noise and the noise-only frames were varied, while the filter size was kept fixed (L^{WF}=20).

Next, the influence of the filter size L^{WF} for each channel was investigated. A size of 5000 samples for the speech-and-noise frame and a size of 15000 samples for the noise-only frame were chosen. The *'start speech'* was taken in the middle of the files, at sample 80000 (=5sec.). The signals were recorded for different locations of the noise source, corresponding to angles between 0^o and 345^o in steps of 15^o (90^o was the side of the hearing aid). The desired speech source was always located at 0^o.

It is important to know if the optimal filtering technique distorts or has an influence on the speech signal itself. To investigate this effect, the transfer function between the system input and output was analyzed with speech signal at 0^o and noise signal at 90^o. For this and the next experiments, the *'start speech'* was taken in the middle of the file. The adaptation of the beamformer second filter was performed during 10000 samples of the noise-only period. With the SVD-based optimal filtering technique, the filter was calculated with a length of 5000 samples for the speech-and-noise frame and 15000 samples for the noise-only frame. The length of the filter for each channel was again twenty.

4.4.1.4 Errors of the VAD

The influence of the VAD errors on the performance of the noise reduction algorithms was examined when the speech source was located at 0^o and a noise source at 90^o. A size of 5000 samples for the speech-and-noise frame and

a size of 15000 samples for the noise-only frame were chosen. For modeling the errors of the VAD, the speech-and-noise frame (F_{SN}) and the noise-only frame (F_N) were redefined as:

$$F_{SN}^{VADerror} = (1 - VAD_{error})\ F_{SN} + VAD_{error}\ F_N \tag{4.34}$$

$$F_N^{VADerror} = (1 - VAD_{error})\ F_N + VAD_{error}\ F_{SN} \tag{4.35}$$

where VAD_{error} denoted the degree of the corruption and when $VAD_{error} = 0$, there was a perfect voice activity detection.

4.4.1.5 Noise scenarios

The experimental set-up to evaluate the performance of the noise reduction strategies in different jammer sound scenarios is presented.

4.4.1.5.1 One noise source

In this experiment, the recordings were made with the same 'ICRA' signals and under the same conditions as described in section 4.4.1.3. The signals were recorded for different locations of the noise source, corresponding to angles between 0^o and 345^o in steps of 15^o.

4.4.1.5.2 Multiple noise sources

For scenarios with multiple and different noise sources, four loudspeakers were used, one for the speech signal (0^o) and three for the noise signals (90^o, 180^o and 270^o). For the speech signal, 'ICRA' signals were used again, but for the noises the ICRA-signal, BLU-signal, NVA-signal and BRUGSE-signal were used. The sum of the signals were recorded with the same level (70dBSPL) at the middle of the dummy head.

4.4.1.5.3 Diffuse noise

Until now, the behaviour with localized noise sources was studied. However, the sound field in a large room or in a car may appear as diffuse noise. A diffuse noise does not have a well-specified direction. To create the diffuse noise, a set-up was made as defined by Veit and Sander (1987).

The reverberation time of the room was $T_{60} = 0.88s$ for a speech-weighted noise. The level of each loudspeaker was fitted to give the same sound level (61dB) at a reference point (middle of eight loudspeakers). With the eight loudspeakers simultaneously, there is a sound level of 70dB at the reference point. For the measurements, the center of the dummy head was placed in the middle of the eight loudspeakers. To create the speech signal, a loudspeaker was put in front of the dummy head at one meter at different angles between 315^o and 45^o in steps of 15^o. The two signal contributions, speech and diffuse noise, were recorded separately.

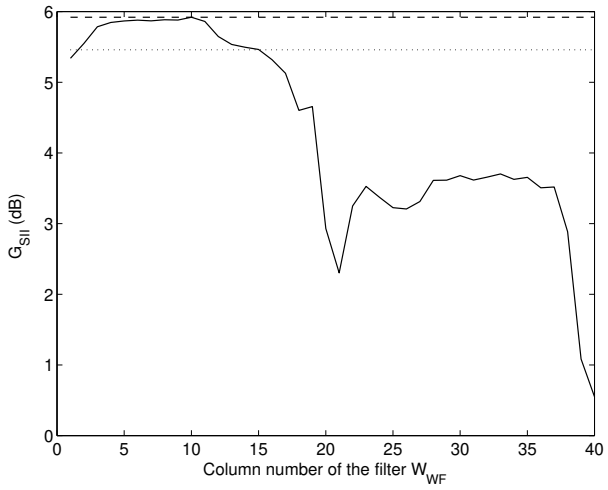


Figure 4.10: Results for the different methods to determine which column of $\mathbf{W}_{WF}$ corresponds to the best estimator. $(-)$ shows the results obtained with the different columns of $\mathbf{W}_{WF,k}$, $(--)$ with the error covariance matrix method and $(\cdots)$ with the average of the estimates.

4.4.2 Results and discussion

4.4.2.1 Optimization of the parameters

The results for the optimization of the SVD-based optimal filtering technique are now presented.

4.4.2.1.1 Optimal estimate

The results are shown in **figure 4.10**. The ECM method clearly and obviously performed better than the averaging method and selects the optimal estimator (column ten). The improvement for each column of $\mathbf{W}_{WF,k}$ showed two different parts, namely between columns one to twenty, and between columns twenty-one to forty. In fact, the first part corresponds to the data of the directional microphone, and the second part corresponds to the data of the omnidirectional microphone. In these two parts, the maximum improvement was reached when the column was taken exactly in the middle, for the two kinds of microphones (omnidirectional and directional). Obviously, there was a better improvement with the directional part because the data of the directional microphone contained less noise than the data of the omnidirectional microphone when the noise source was located at 90^o (remember that $SNR_{SII,input}$ in formula (2.6) corresponds to the front omnidirectional micro-

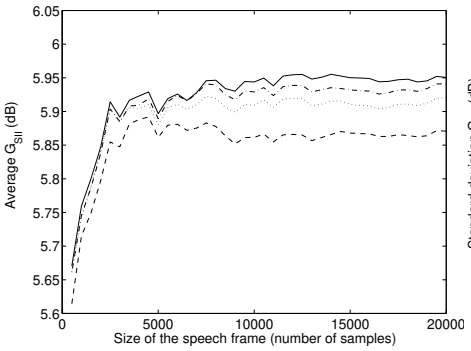


Figure 4.11: Influence of the length of the noise-only and the speech-and-noise frames on the mean of the G_{SII} improvement. The size of the filters for one channel is $L^{WF} = 20$. $(--)$ shows the results obtained with a size of the noise-only frame=5000 samples, $(\cdots)$ with a size of the noise-only frame=10000 samples, $(-\cdot)$ with a size of the noise-only frame=15000 samples and $(-)$ with a size of the noise-only frame=20000samples.

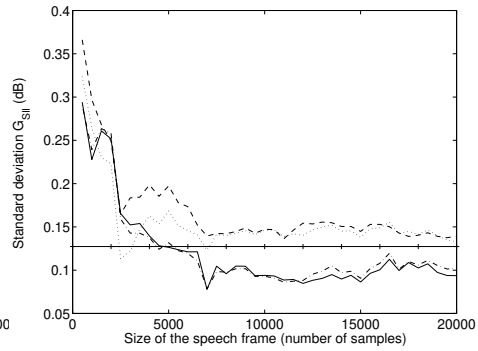


Figure 4.12: Influence of the length of the noise-only and the speech-and-noise frames on the standard deviation of the G_{SII} improvement. The size of the filters for one channel is $L^{WF} = 20$. $(--)$ shows the results obtained with a size of the noise-only frame=5000 samples, $(\cdots)$ with a size of the noise-only frame=10000 samples, $(-\cdot)$ with a size of the noise-only frame=15000 samples and $(-)$ with a size of the noise-only frame=20000 samples, $(++)$ shows the standard deviation obtained with the two-stage adaptive beamformer.

phone). The ECM method showed that column ten ($L^{WF}/2$) gave the best result (for $M = 2$), and this agrees with the result obtained by Doclo (2002). This is certainly due to the fact that the middle column of $\mathbf{W}_{WF,k}$ has symmetric properties (for L^{WF} is odd) and results in a linear phase filter (Doclo, 2003). The averaging filter ($\mathbf{w}_2$) cannot be an optimal filter because $\mathbf{w}_2$ is obtained averaging the collection of L^{WF}-dimensional optimal filters. From now on (and for the sake of simplicity) column $L^{WF}/2$ of the $\mathbf{W}_{WF,k}$ will always be taken.

4.4.2.1.2 Influence of the design parameters

Figures 4.11 and **4.12** show the effect of the size of the noise-only frame and the speech-and-noise frame on the mean and the standard deviation of the G_{SII} improvement. The values of the beamformer were taken as references, i.e. 6.32dB for the mean and 0.13dB for the standard deviation. In the case of a large standard deviation, it was found that the signal had fluctuations which can correspond to unnatural sounds. The two-stage adaptive beamformer with this standard deviation (0.13dB) was know to give satisfactory results in this

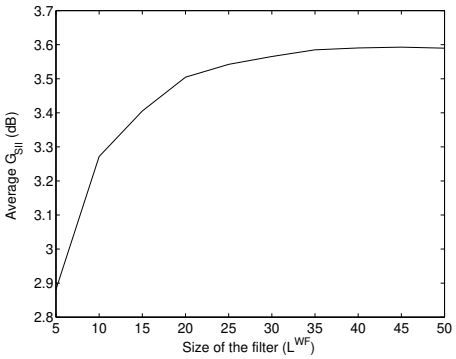


Figure 4.13: Influence of the filter size for each channel on the average G_{SII} improvement. The length of the speech-and-noise frame is 5000, the length of the noise-only frame is 15000.

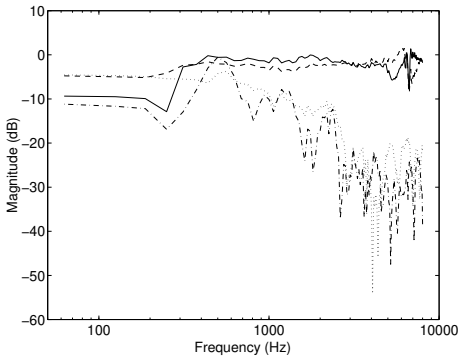


Figure 4.14: Magnitude of the transfer function estimate between the directional microphone and the output of the algorithms for speech-only and for noise-only. $(-)$ shows the transfer function of the two-stage adaptive beamformer for speech and $(-\cdot)$ for noise. $(--)$ shows the transfer function of the SVD-based technique for speech and $(\cdots)$ for noise.

respect.

Up to 5000 samples, if the size of the speech-and-noise frame were increased, the G_{SII} improvement increased and the standard deviation dropped. After 5000 samples, there was only a marginal improvement. For the noise-only frame, when the length was increased, the same effect as with the speech-and-noise frame was obtained. The G_{SII} improvement increased and the standard deviation decreased. So the longer the noise-only frame, the larger the G_{SII} improvement. To obtain the same performance as the beamformer (with adaptation of the second filter on 10000 samples) for the standard deviation, a noise-only frame longer than 10000 samples was taken (15000 samples). Furthermore, for the size of frame that was used, the SVD-based algorithm performed slightly worse (0.4dB) than the beamformer.

Figure 4.13 shows the average G_{SII} over all azimuth angles (between 0^o and 345^o) in function of the filter size L^{WF} for each channel. For a length $L^{WF} > 20$, the average G_{SII} improvement stayed roughly the same, while for $L^{WF} < 20$ the average G_{SII} was seen to drop significantly.

Figure 4.14 shows the transfer function between the directional microphone and the output of the two algorithms. The optimal filtering technique performed as well as the beamformer and the transfer function for speech was

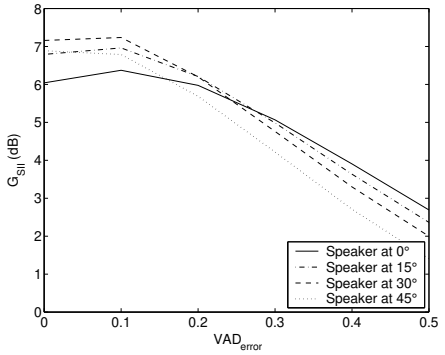


Figure 4.15: Influence of VAD errors on the G_{SII} for the two-stage adaptive beamformer.

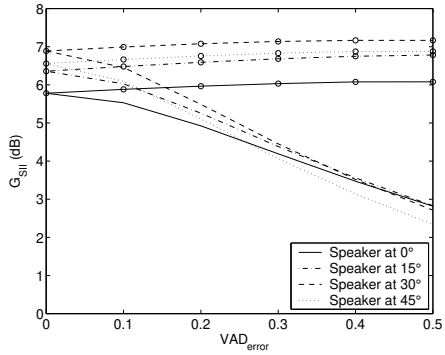


Figure 4.16: Influence of VAD errors on the G_{SII} for the optimal filtering technique when the noise statistics are corrupted by the speech statistics (−) and when the speech statistics are corrupted by the noise statistics (o).

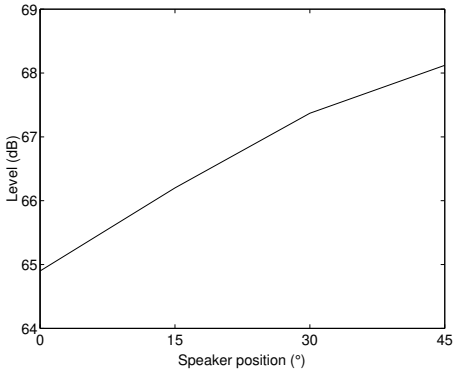


Figure 4.17: Level (dB) of the speech signal in the noise reference of the A2B as a function of the speaker position.

around 0dB for the frequencies between 300Hz and 6.5kHz. The attenuation at the low frequencies (below 500Hz) was due to the DC filter. Thus, it can be concluded that both noise reduction algorithms do not cause any distortion of the speech signal.

From these experiments, it can be concluded that the SVD-based technique roughly performs as well as the two-stage adaptive beamformer. Indeed, with a filter size of 2×20 coefficients for the optimal filtering technique and $10 + 30$ coefficients for the beamformer, the speech intelligibility improvement

was only 0.4dB higher for the beamformer. Furthermore, this improvement could be obtained with a similar adaptation time for the coefficients (10000 samples) for both algorithms. However, to obtain a similar standard deviation in the G_{SII} a slightly longer adaptation time was needed for the SVD-based technique.

4.4.2.2 Errors of the VAD

Figures 4.15 and **4.16** show the influence of the VAD errors as a function of the speaker position on the performance of the two-stage adaptive beamformer and the optimal filtering technique, respectively.

Because the hearing aid was worn by a mannequin during the experiments, the SNR at the microphones of the device depends on the speaker positions resulting in different values of the G_{SII} for a perfect VAD ($VAD_{error} = 0$). Although the first filter of the adaptive beamformer was calibrated for the hearing aid, the strategy was affected by the VAD errors when the speech source was at 0^o. This means that the first filter of the A2B was calibrated in anechoic conditions and failed in reverberant conditions allowing a leakage of the speech signal in the noise reference of the strategy (**figure 4.17**).

The adaptive filter of the A2B was corrupted by the speech statistics and brought a cancellation of the speech signal at the output of the A2B. This was predicted by the theoretical evaluation when the experiment combined a gain deviation on the rear microphone with VAD errors (**figure 4.7**). The higher the level of the speech signal in the noise reference, the more severe the cancellation of the speech signal at the output of the A2B (**figure 4.15**). As predicted by the theoretical evaluation, the optimal filtering technique was affected by the VAD errors when the statistics of the noise signal were corrupted by the speech signal but not when the statistics of the speech-and-noise signal were corrupted by the noise-only signal. However, if the statistics of the noise signal are not updated during a long time before a speech-and-noise period, the optimal filtering technique cannot perform an optimal noise reduction especially in non-stationary conditions.

4.4.2.3 Behaviour with different noise scenarios

The noise reduction performance of both strategies are examined in different noise scenarios when a perfect VAD is assumed ($VAD_{error} = 0$).

4.4.2.3.1 One noise source

Figure 4.18 shows the behaviour of the omnidirectional microphone, the software directional microphone, the output of the two-stage adaptive beamformer and the output of the optimal filtering technique. Ideally, the omnidi-

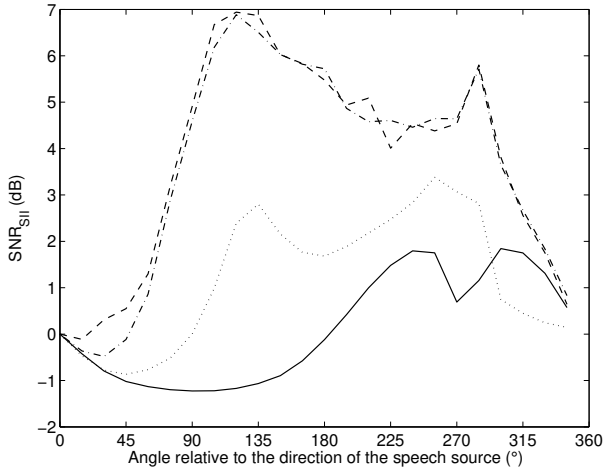


Figure 4.18: The curves show the SNR_{SII} at the omnidirectional microphone $(-)$ as a function of angle relative to the direction of the speech source, at the directional microphone $(\cdots)$, at the output of the two-stage adaptive beamformer $(--)$ and at the output of the SVD-based technique $(-\cdot)$.

Table 4.1: G_{SII} improvement (in dB) for multiple noise sources.

			A2B	SVD	Dir
N_{90}	N_{180}	N_{270}	G_{SII}	G_{SII}	G_{SII}
ICRA			6.0	5.7	1.2
	ICRA		5.6	5.7	1.7
		ICRA	3.8	3.9	3.0
ICRA	BLU		3.8	3.8	2.0
ICRA		BLU	5.6	5.3	2.2
NVA	BLU		4.1	4.0	1.7
NVA		BLU	6.0	5.3	2.1
ICRA	ICRA		2.8	2.8	1.3
ICRA		ICRA	4.9	4.7	2.2
ICRA	BLU	NVA	3.3	3.4	2.2
NVA	BLU	BRUG	4.0	3.9	2.0
ICRA	ICRA	ICRA	2.8	2.9	1.9

rectional microphone has the same sensitivity for all angles (around 0dB in our case). However, the sensitivity was seen to be a function of the angle, mainly due to the effect of the head. **Figure 4.18** shows that the beamformer and the SVD-based procedure behave very similarly, and so the improvement is the same for the two algorithms. These algorithms always give a larger improvement than the omnidirectional and the directional microphone. Remarkably,

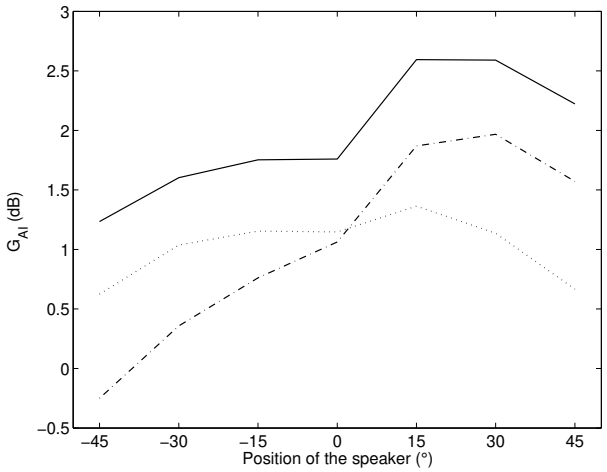


Figure 4.19: In the diffuse noise case, the G_{AI} improvement between the omnidirectional microphone and the directional microphone ($\cdots$), the two-stage adaptive beamformer ($-\cdot$) and the optimal filtering technique ($-$).

the directional microphone can perform worse than the omnidirectional microphone, namely between angles 285^o and 345^o.

4.4.2.3.2 Multiple noise sources

For scenarios with multiple and different noise sources, the results are shown in **table 4.1**. Evidently, with two and three noise sources, both the beamformer and the SVD-based algorithm performed worse than in the one noise source case. These two strategies give better results than the directional microphone and had roughly the same G_{SII} improvement for the different situations. As in the one noise source configuration, the head diffraction contributed to the improvement. Indeed, the G_{SII} improvement was larger when there was a noise source at 270^o instead of 180^o. Also, a better G_{SII} improvement could be observed when the noise source did not contain an ICRA-signal spectrum.

4.4.2.3.3 Diffuse noise

In the diffuse noise case, the room was also highly reverberant and hence the G_{SII} procedure could not be applied properly. For this reason the G_{AI} was used, as introduced in the section 2.2.1.1. The G_{AI} improvement of the optimal filtering technique was always better than the directional microphone and the two-stage adaptive beamformer (**figure 4.19**). For angles below 0^o the A2B performed worse than the directional microphone and the omnidirectional

microphone, in the case where the speaker was at -45^o. The difference between the directional microphone and the SVD-based optimal filtering technique was small but important for hearing-aid users.

4.5 Real time implementation

In this section, an extension of the SVD-based optimal filtering technique is introduced and a real time implementation of the strategy is described.

4.5.1 Speech distortion parameter

A speech distortion parameter μ is introduced in the SVD-based optimal filtering technique to minimize the energy of the signal distortion ϵ_s^2 under the constraint that the residual noise energy ϵ_n^2 stays under a threshold T_n (section 2.2.3.2.4) (Ephraim and Van Trees, 1995).

$$\min_{\mathbf{W}_k} \quad \epsilon_s^2 \quad subject \quad to \quad \epsilon_n^2 \leq T_n \quad where \quad 0 \leq T_n \leq 1 \tag{4.36}$$

The filter $\mathbf{W}_{WF}$ of the equation (4.30) becomes:

$$\mathbf{W}_{WF} = \mathbf{X}^{-T} \; diag\left\{ \frac{q\ \sigma_i^2 - p\ \eta_i^2}{q\ \sigma_i^2 + (\mu - 1)p\ \eta_i^2} \right\} \mathbf{X}^T \tag{4.37}$$

where the speech distortion parameter $\mu \in [0, \infty[$ allows a trade-off between signal distortion and noise reduction. If $\mu = 1$ the MMSE solution is obtained. More emphasis is put on the signal distortion when $\mu < 1$ at the expense of decreasing the noise reduction performance. The residual noise level is reduced when $\mu > 1$ at the expense of increasing the speech distortion. With $\mu \to \infty$, all the emphasis is put on the noise reduction without taking into account the signal distortion.

4.5.2 Real time implementation

The real time implementation of the SVD-based technique is illustrated in **figure 4.20**. Five steps are necessary to compute the filter coefficients in real time:

• *Step* 1 : The VAD used in the real time implementation of the optimal filtering technique is the same as the VAD used in the real time implementation of the two-stage adaptive beamformer (section 3.3.1.3).

• *Step* 2 : Classification errors between the speech-and-noise periods and the noise-only periods occur with the VAD. If the speech-and-noise periods are wrongly classified, speech-and-noise vectors are added to the noise

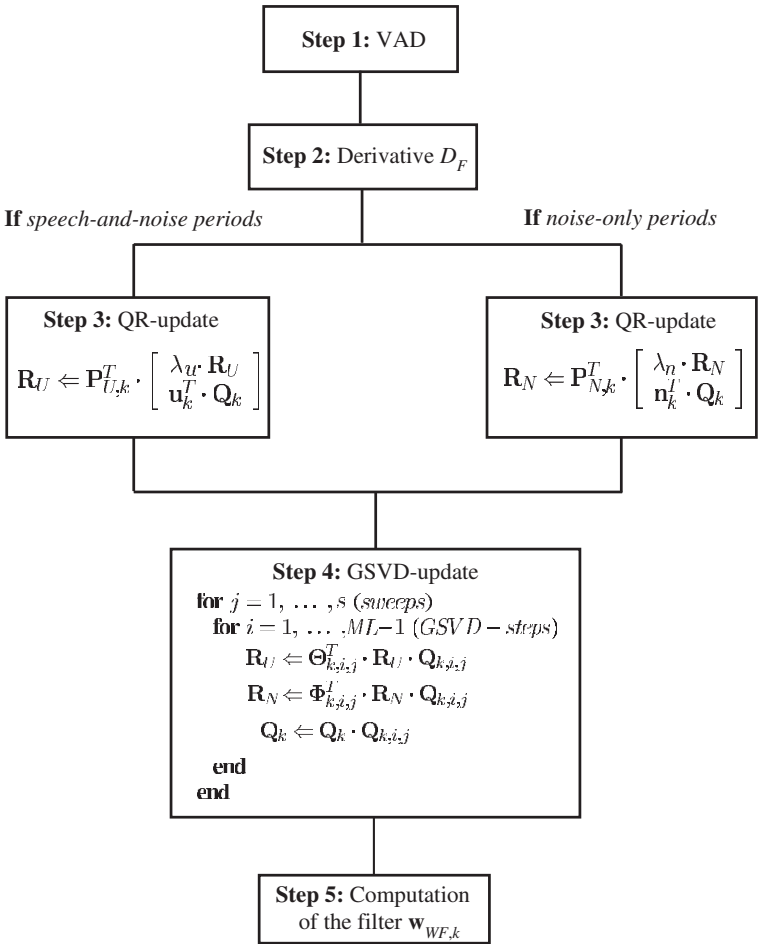


Figure 4.20: Real time implementation of the SVD-based optimal filtering technique.

matrix ($\mathbf{N}_k$). In this case, the factor $F \triangleq \sum_{i=1}^{2L^{WF}} 1 - \eta_i^2/\sigma_i^2$ of the filter $\mathbf{W}_{WF}$ tends to be small ($\sigma_i^2 \rightarrow \eta_i^2$), resulting in a zeroing of the filter coefficients of the optimal filtering technique. Since F varies in time, the derivative D_F of this factor can be measured during the processing:

$$D_F = \frac{\delta[1/(2L^{WF}) \ \sum_{i=1}^{2L^{WF}} (1 - \eta_i^2/\sigma_i^2)]}{\delta t} \tag{4.38}$$

If the derivative D_F is below a given threshold T_D, this can mean that the VAD misclassified speech-and-noise periods. Then, a correction is brought

to the VAD and the decision made in *Step 1* is modified. Otherwise, when $D_F > T_D$, the decision made in *Step* 1 is kept valid.

The behaviour of the derivative D_F for a perfect and a non-perfect VAD is illustrated **figure 4.21**. **Figure 4.21a** shows a noisy speech signal with a SNR of 0dB when a speech source was present at 0^o and a stationary speech weighted noise source at 90^o. For a perfect VAD (**figure 4.21b**), the behaviour of the factor F is illustrated in **figure 4.21c**. After convergence, the distance between the generalized singular values η_i and σ_i was constant and the factor F did not vary over time, resulting in a small D_F ($D_F \to 0$) (**figure 4.21d**). For a non perfect VAD (**figure 4.21e**), the behaviour of the factor F is illustrated in **figure 4.21f**. Abrupt decreases in F could be observed when VAD errors occurred because the generalized singular values η_i of the matrix $\mathbf{N}_k$ converged to the generalized singular values σ_i of the matrix $\mathbf{U}_k$. In this case, the factor F decreased as a function of time, resulting in negative values of the derivative D_F **figure 4.21g**. During the initialization of the optimal filtering technique (first three seconds of the **figure 4.21g**), the matrix $\mathbf{U}_k$ was only updated by noise-only data and $D_F \to -\infty$.

The function D_F is effective ($D_F < T_D$) with VAD errors, but can also be effective with temporal variation of a non-stationary noise. For instance, in a multitalker babble (perfect VAD is assumed) the level of the generalized singular value η_i varies while the level of the generalized singular value σ_i stays roughly the same. When the generalized singular value η_i increase, the factor F becomes smaller resulting in the decrease of the derivative D_F. If the latter function becomes effective ($D_F < T_D$), the noise signal is detected as a speech-and-noise signal and the optimal filtering technique does not perform an optimal cancellation of the noise source. In this study, the threshold T_D was chosen low enough to avoid that the function D_F was effective ($D_F < T_D$) with a perfect VAD in stationary and non-stationary (e.g. multitalker babble) noise at a SNR of -10dB. The threshold T_D was also chosen high enough so that no temporal distortions were observed when the VAD could only detect correctly 50% of the speech-and-noise periods. In the sequel, the value of T_D was -0.00015.

• *Step* 3 : A recursive technique was introduced by Doclo and Moonen (2002) to approximate the SVD-based optimal filtering strategy. The technique is based on a Jacobi-type GSVD-updating algorithm (Moonen et al., 1991). Recursive GSVD-updating algorithms use the decomposition of the GSVD at time $k-1$ to compute the decomposition at time k. The equation 4.29 at time $k-1$ can be rewritten as:

$$\begin{cases} \mathbf{U}_{k-1} = \mathbf{Y}_{k-1}\ \mathbf{R}_{U,k-1}\ \mathbf{Q}_{k-1}^T \\ \mathbf{N}_{k-1} = \mathbf{V}_{k-1}\ \mathbf{R}_{N,k-1}\ \mathbf{Q}_{k-1}^T, \end{cases} \tag{4.39}$$

where $\mathbf{R}_{U,k-1} \in \mathbb{R}^{ML^{WF} \times ML^{WF}}$ and $\mathbf{R}_{N,k-1} \in \mathbb{R}^{ML^{WF} \times ML^{WF}}$ are upper tri-

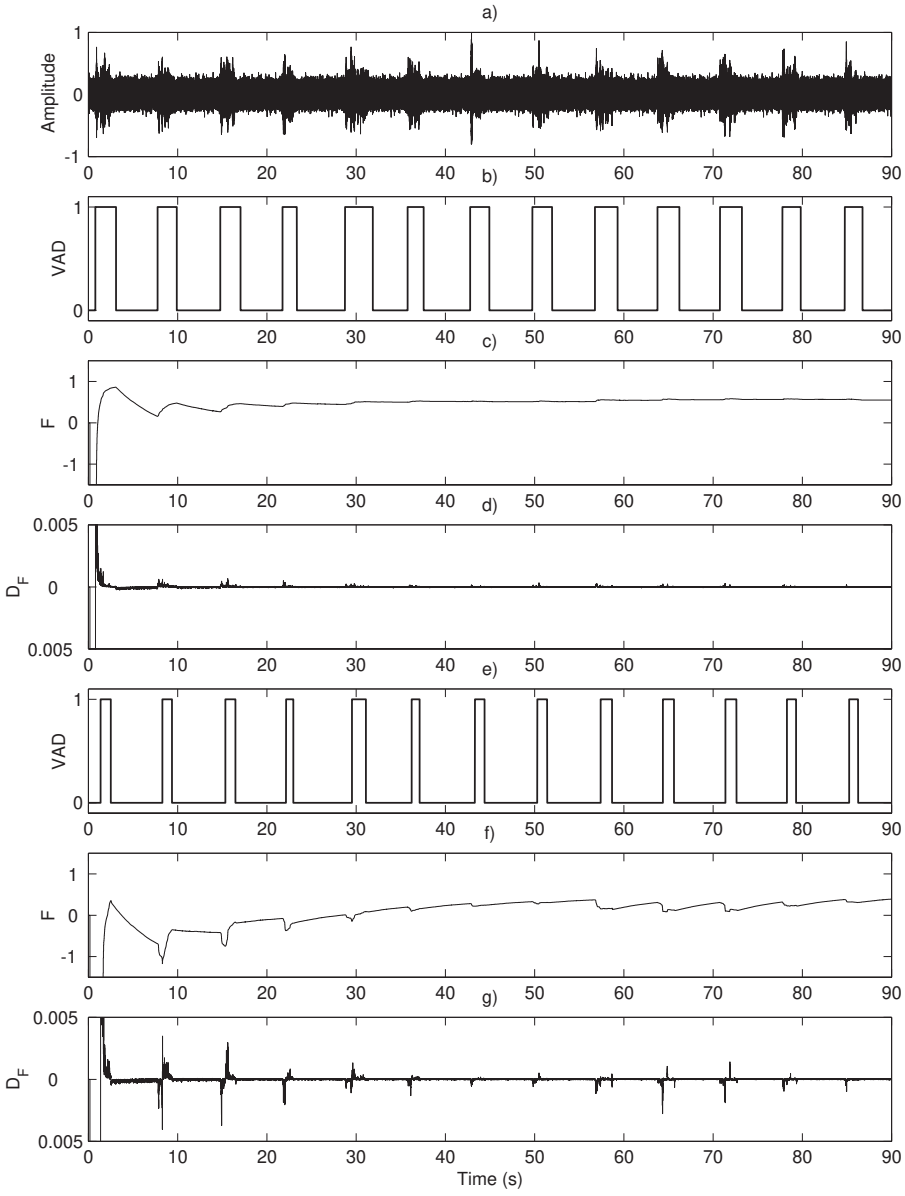


Figure 4.21: Behaviour of the derivative D_F as a function of the time. a) A speech source was at 0^o and a noise source at 90^o with a SNR of 0dB, b) with a perfect VAD, c) the behaviour of the factor F, d) and the derivative D_F is shown, e) with an unperfect VAD f), the behaviour of the factor F, g) and the derivative D_F is shown.

angular matrices having parallel rows and $\mathbf{Q}_{k-1} \in \mathbb{R}^{ML^{WF} \times ML^{WF}}$ is an orthogonal matrix. For the recursive computation, only $\mathbf{R}_{U,k-1}$, $\mathbf{R}_{N,k-1}$ and $\mathbf{Q}_{k-1}$ have to be stored. When a new data vector $\mathbf{u}_k$ (speech-and-noise) or $\mathbf{n}_k$ (noise-only) is present, a new update needs to be performed such that $\mathbf{U}_k$ and $\mathbf{N}_k$ are defined in a recursive manner:

$$\mathbf{U}_k = \begin{bmatrix} \lambda_u \ \mathbf{U}_{k-1} \\ \mathbf{u}_k \end{bmatrix} \quad or \quad \mathbf{N}_k = \begin{bmatrix} \lambda_n \ \mathbf{N}_{k-1} \\ \mathbf{n}_k \end{bmatrix} \tag{4.40}$$

where λ_u and λ_n are exponential weighting factors for speech-and-noise and noise-only matrix respectively. For the general case $\mathbf{U}_k$ and $\mathbf{N}_k$ can be rewritten as

$$\mathbf{U}_k = \begin{bmatrix} \mathbf{Y}_{k-1} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ 0 \cdots 0 & 1 \end{bmatrix} \begin{bmatrix} \lambda_u \ \mathbf{R}_{U,k-1} \\ \mathbf{u}_k^T \ \mathbf{Q}_{k-1} \end{bmatrix} \mathbf{Q}_{k-1}^T \tag{4.41}$$

$$\mathbf{N}_k = \begin{bmatrix} \mathbf{V}_{k-1} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ 0 \cdots 0 & 1 \end{bmatrix} \begin{bmatrix} \lambda_n \ \mathbf{R}_{N,k-1} \\ \mathbf{n}_k^T \ \mathbf{Q}_{k-1} \end{bmatrix} \mathbf{Q}_{k-1}^T \tag{4.42}$$

The triangular factors are thus restored by QR-updates with the transformed input vectors $\tilde{\mathbf{u}}_k^T = \mathbf{u}_k^T \ \mathbf{Q}_{k-1}$ or $\tilde{\mathbf{n}}_k^T = \mathbf{n}_k^T \ \mathbf{Q}_{k-1}$. QR-updating is performed by using orthogonal Givens rotations, zeroing the elements on the bottom row of the component matrices with $\mathbf{u}_k$ (or $\mathbf{n}_k$) in (4.41) (or (4.42)). When a new data vector is present, only one QR-update is carried out. Indeed, if a noise-only period is present, a QR-update for $\mathbf{N}_k$ is performed and if a speech-and-noise period is present, a QR-update for $\mathbf{U}_k$ is performed. Assuming that speech-and-noise period is present, the QR-update for $\mathbf{U}_k$ is written as the equation (4.43), where $\mathbf{P}_{U,k}$ is an $(ML^{WF}+1) \times ML^{WF}$ matrix with orthogonal columns, which does not need to be computed explicitly, and $\tilde{\mathbf{R}}_{U,k}$ the updated triangular matrix. The matrix $\mathbf{Q}_{k-1}$ is not altered by the QR-update.

$$\mathbf{U}_k = \underbrace{\begin{bmatrix} \mathbf{Y}_{k-1} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ 0 \cdots 0 & 1 \end{bmatrix}}_{\tilde{\mathbf{Y}}_k} \mathbf{P}_{U,k} \ \tilde{\mathbf{R}}_{U,k} \ \mathbf{Q}_{k-1}^T \tag{4.43}$$

$$\mathbf{N}_k = \underbrace{\begin{bmatrix} \mathbf{V}_{k-1} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ 0 \cdots 0 & 1 \end{bmatrix}}_{\tilde{\mathbf{V}}_k} \mathbf{P}_{N,k} \ \tilde{\mathbf{R}}_{N,k} \ \mathbf{Q}_{k-1}^T \tag{4.44}$$

• *Step* 4 : After each QR-update the GSVD process is resumed. The GSVD of $\tilde{\mathbf{R}}_{U,k}$ and $\tilde{\mathbf{R}}_{N,k}$ is computed by carrying out an iterative procedure, where a series of Givens transformations is applied to $\tilde{\mathbf{R}}_{U,k}$ and $\tilde{\mathbf{R}}_{N,k}$ in order to yield upper triangular factors with parallel rows (Moonen et al., 1991). In **figure 4.20**, the matrices $\mathbf{Q}_{k,i,j}$, $\mathbf{\Theta}_{k,i,j}$ and $\mathbf{\Phi}_{k,i,j}$ correspond to the Givens rotations. For details on the updating scheme, the reader is referred to (Moonen et al., 1991; Doclo, 2003).

• *Step* 5 : This step consists of computing the optimal filter $\mathbf{w}_{WF,k}$ after the update of the recursive GSVD-updating algorithm. Substituting formulae (4.39) into (4.10), the equation can be rewritten as:

$$\mathbf{W}_{WF,k} = \mathbf{Q}_k \mathbf{R}_{U,k}^{-1} diag\left\{ \frac{(1-\lambda_n^2)(\mathbf{R}_{U,k}^{ii})^2 - (1-\lambda_u^2)(\mathbf{R}_{N,k}^{ii})^2}{(1-\lambda_n^2)(\mathbf{R}_{U,k}^{ii})^2 + (\mu-1)(1-\lambda_u^2)(\mathbf{R}_{N,k}^{ii})^2} \right\} \mathbf{R}_{U,k} \mathbf{Q}_k^T \tag{4.45}$$

The factor p/q is replaced by $(1-\lambda_n^2)/(1-\lambda_u^2)$. Since only one column (the $i-th$ column, $\mathbf{w}_{WF,k}^i$ of $\mathbf{W}_{WF,k}$) needs to be computed, this column can be computed as the solution of the linear set of equations:

$$\mathbf{R}_{U,k} \underbrace{\mathbf{Q}_k^T \mathbf{w}_{WF,k}^i}_{\tilde{\mathbf{w}}_k} = \underbrace{diag\left\{ \frac{(1-\lambda_n^2)(\mathbf{R}_{U,k}^{i,i})^2 - (1-\lambda_u^2)(\mathbf{R}_{N,k}^{i,i})^2}{(1-\lambda_n^2)(\mathbf{R}_{U,k}^{i,i})^2 + (\mu-1)(1-\lambda_u^2)(\mathbf{R}_{N,k}^{i,i})^2} \right\} \mathbf{R}_{U,k} \mathbf{Q}_k^{i\,T}}_{\tilde{\mathbf{x}}_k} \tag{4.46}$$

The calculation of the $\mathbf{w}_{WF,k}^i$ consists of computing $\tilde{\mathbf{x}}_k$, solving the equation $\mathbf{R}_{U,k}\ \tilde{\mathbf{w}}_k = \tilde{\mathbf{x}}_k$ by back-substitution and computing $\mathbf{w}_{WF,k}^i$ as:

$$\mathbf{w}_{WF,k}^i = \mathbf{Q}_k\ \tilde{\mathbf{w}}_k \tag{4.47}$$

The number of filter-taps was optimized for the two-stage adaptive beamformer in the section (3.3.2.2) and for the SVD-based optimal filtering technique in the section (4.4.2.1.2). At 16kHz, the number of the filter-taps were ten and thirty for the first and the second filter of the adaptive beamformer respectively, and the number of filter-taps was twenty for the optimal filtering technique. In the sequel, the noise reduction strategies are implemented in real time at 12kHz and the number of filter taps are redefined in order to keep the ratio of the filter length to sampling frequency fixed. The length of the first and the second filter of the adaptive beamformer are then eight and twenty two respectively, and the length of the filters per channel for the optimal filtering technique is fifteen.

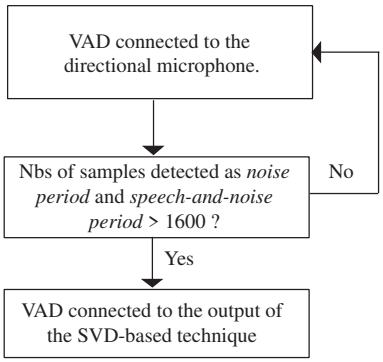


Figure 4.22: Initialization strategy when the VAD is connected to the output of the SVD-based noise reduction scheme.

4.6 Improvement of the VAD performance

The theoretical and the physical experiments demonstrated that VAD errors may affect severely the performance of the noise reduction strategies. The VADs discriminate well the speech-and-noise periods from the noise-only periods under high SNR and stationary noise conditions. In the case of low SNR and non-stationary noise, the performance drops significantly (Tanyer and Ozer, 2000). In this section, a method to improve the performance of VAD in low SNR is described. The publications related to this section are (Maj et al., 2002c; Maj et al., 2003c).

4.6.1 Method

In noise reduction strategies, several signals may be available to the VAD, such as the signal of the omnidirectional microphone, the directional microphones (software or hardware) or the output of the noise reduction technique. In this study, the behaviour of the VAD was evaluated when the VAD was connected to these different signals. The signals of the omnidirectional or the directional microphone (software or hardware) are directly available. However, when the VAD is connected to the output of the noise reduction algorithms, the signals are only available after a first update of the adaptive filters. The optimal filtering technique needs at least a noise-only period and a speech-and-noise period. To solve this problem of initialization , the VAD is initially connected to the directional microphone and when several samples (e.g. 1600) are classified as speech-and-noise periods and noise-only periods, the optimal filters are updated (**figure 4.22**). Only then, the VAD algorithm

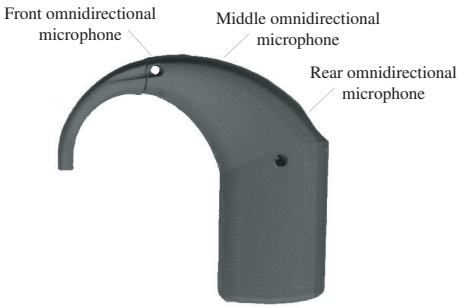


Figure 4.23: Nucleus prototype hearing aid (OO) with three omnidirectional microphones mounted in an endfire array configuration.

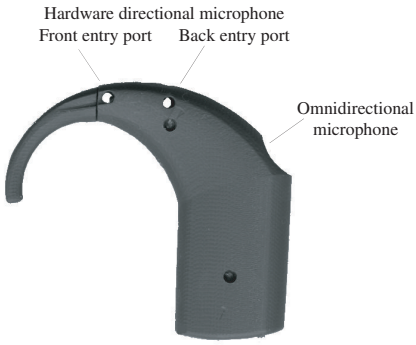


Figure 4.24: Nucleus prototype hearing aid (DO) with one hardware directional microphone and one omnidirectional microphone mounted in an endfire array configuration.

is connected to the output of the SVD-based optimal filtering strategy.

4.6.2 Experimental set-up

For the experiments, the same room as for the large scale study of the two-stage adaptive beamformer was used (section 3.4.1).

4.6.3 Hearing aids

To evaluate the performance of the noise reduction algorithms, two types of BTE hearing aids were used and referred to as DO and OO. They had two different microphone configurations and were prototypes based on a Nucleus BTE headset housing for cochlear implants of Cochlear Ltd. The first type OO (**figure 4.23**) had three omnidirectional microphones (Knowles FG-3452) mounted in an endfire configuration. In this study, only the front and the middle microphones, spaced 1cm apart, were used. A software directional microphone was implemented and the delay operation on the rear microphone was chosen to give a hypercardioid spatial characteristic in anechoic conditions (null at 110^o). The software directional microphone and the rear omnidirectional microphone signal are then used as the inputs to the noise reduction algorithms. The second type DO (**figure 4.24**) had one hardware directional microphone (Microtronic 6001), as front microphone, and one omnidirectional microphone (Knowles FG-3452), as rear microphone. The hardware directional microphone had a hypercardioid spatial characteristic (null at 110^o) in anechoic conditions for a speech weighted noise. The distance between the front entry port and the

back entry port of the hardware directional microphone was 1cm. The distance between the front entry port of the hardware directional microphone and the omnidirectional microphone was 2.5cm.

4.6.3.1 Signals

In the test room, the hearing aid was positioned on a dummy head and two loudspeakers were used. One loudspeaker was located in front of the dummy head (at 0^o) and presented the speech material (sentences), and a second at 90^o (on the side of the hearing aid) presented the noise material. The VAD was evaluated for stationary and non-stationary noise. As stationary noise, a long term average speech-weighted noise of sentences was used, and as non-stationary noise, a multitalker noise was used. The signals from the two omnidirectional microphones of the hearing aid were digitized at a sampling frequency of 12kHz on a PC-platform. The signals were recorded during ninety seconds. The speech and the noise signals were recorded separately. In this way, it was feasible to know perfectly the speech-and-noise periods and the noise-only periods. This perfect VAD was used as a reference for the evaluation of the VAD algorithm.

4.6.3.2 Performance measure

The performance of the VAD was evaluated by calculating the percentage of correctly detected samples by the VAD algorithm for the speech-and-noise periods and noise-only periods of the signals. The percentage (Per) was calculated by:

$$Per = \frac{SN_{RealTime} \times 100}{SN_{Perfect}} \quad or \quad Per = \frac{N_{RealTime} \times 100}{N_{Perfect}} \tag{4.48}$$

where $N_{Perfect}$ and $SN_{Perfect}$ are the number of samples, which are known to be belong to noise-only periods (N) or speech-and-noise periods (SN), based on the 'perfect' VAD. $N_{RealTime}$ and $SN_{RealTime}$ are the number of samples which are correctly classified as noise-only periods or speech-and-noise periods by the real time VAD. In the calculation, the first twenty seconds of the signals were not taken into account, which gave time to the noise reduction algorithms to converge.

4.6.4 Results

4.6.4.1 Stationary conditions

Figures 4.25 and **4.26** show the results of the percentage (Per) of samples correctly detected by the VAD algorithm in a stationary speech weighted noise by the VAD algorithm for speech-and-noise periods and noise-only periods. For both types of the hearing aids, OO and DO, the VAD algorithm

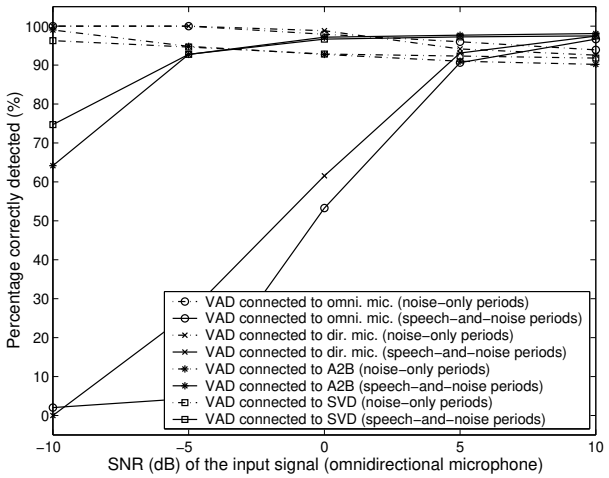


Figure 4.25: Performance of the VAD when it is connected to the front omnidirectional microphone, the software directional microphone, the output of the adaptive beamformer and the output of the SVD-based technique in stationary conditions (hearing aid OO)

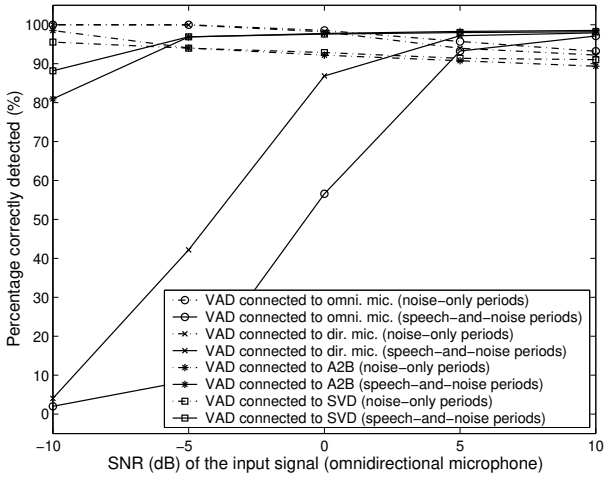


Figure 4.26: Performance of the VAD when it is connected to the front omnidirectional microphone, the hardware directional microphone, the output of the adaptive beamformer and the output of the SVD-based technique in stationary conditions (hearing aid DO)

detected almost correctly ($Per > 90\%$) the noise-only periods when it was connected to the omnidirectional microphone, the directional microphone (software and hardware) or the output of the noise reduction strategies (two-stage adaptive beamformer or optimal filtering technique). The detection performance for the speech-and-noise periods was clearly dependent on where the VAD was connected. The performance of the VAD dropped significantly when it was connected to the omnidirectional microphone or the directional microphones for a SNR below 5dB. When the VAD was connected to the output signal of the noise reduction algorithm, the percentage of well-detected samples stayed above 90% for a SNR above -5dB. With the hearing aid OO at SNR=-10dB, the percentage obtained with the optimal filtering technique was 75% and with the two-stage adaptive beamformer the percentage was 65%. With the hearing aid DO at SNR=-10dB, the scores were about 90% with the optimal filtering technique and about 80% with the adaptive beamformer. Per was roughly similar for the hearing aids with the two omnidirectional microphones. The hardware directional microphone gave a better performance than the software directional microphone for a SNR below 0dB.

4.6.4.2 Non-stationary conditions

In non-stationary conditions, the performance of the VAD is illustrated in **figures 4.27** and **4.28**. The omnidirectional microphone of the device DO had a higher Per than the omnidirectional microphone of the device OO for the speech-and-noise periods below 5dB SNR. The percentages for the noise-only periods were similar. This may be due to the fact that the omnidirectional microphones were not at the same position in the hearing aids.

Connecting the VAD to the software and the hardware directional microphones did not have the same impact on the detection performance of the VAD. With the software directional microphone, the noise-only periods were well-detected ($Per > 90\%$) independent of the SNR, while detection of the speech-and-noise periods was dependent on the SNR. The lower the SNR, the lower the Per. Relative to the software directional microphone, the VAD connected to the hardware directional microphone detected better the speech-and-noise periods but performed worse for the noise-only periods ($Per \approx 45\%$) relative to 85%. The VAD-output is a function of the statistical variations of the signals and a speech-and-noise period is detected if the statistics of the new frame exceed the estimated statistics of the previous frames. The more the statistics change, the more the speech-and-noise periods are detected. **Figure 4.29** shows the frequency responses of the hardware directional microphone and the software directional microphone. The directional processes attenuate the sensitivity at the low frequencies ($f < 1$kHz). Both directional microphones have roughly the same frequency response (intermicrophone or interport distance in both cases is 10mm). Between the frequencies 1kHz and 4kHz, the frequency response of the hardware directional microphone exceeds the fre-

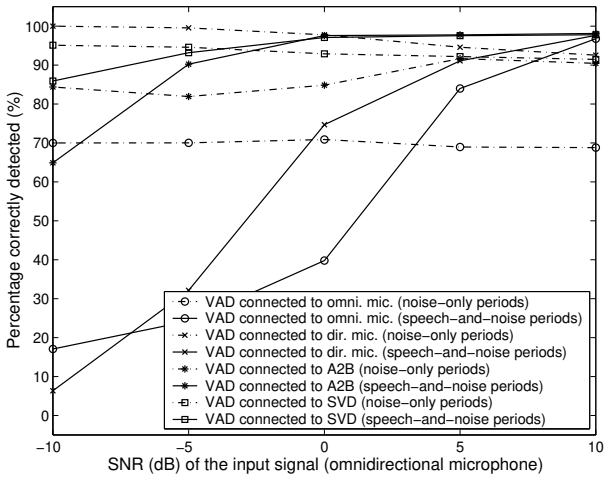


Figure 4.27: Performance of the VAD when it is connected to the front omnidirectional microphone, the software directional microphone, the output of the adaptive beamformer and the output of the SVD-based technique in non-stationary conditions (hearing aid *OO*)

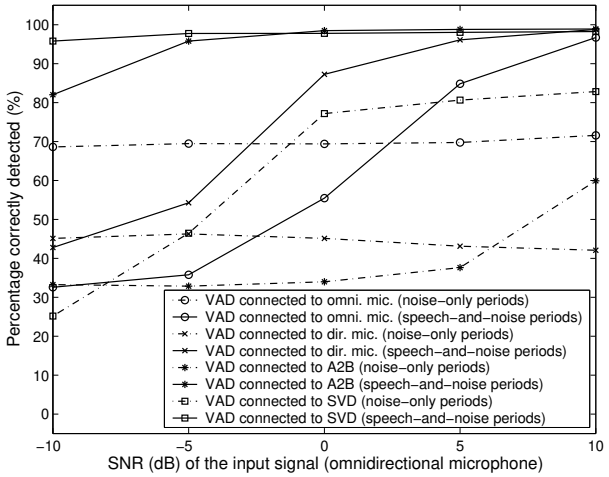


Figure 4.28: Performance of the VAD when it is connected to the front omnidirectional microphone, the hardware directional microphone, the output of the adaptive beamformer and the output of the SVD-based technique in non-stationary conditions (hearing aid *DO*)

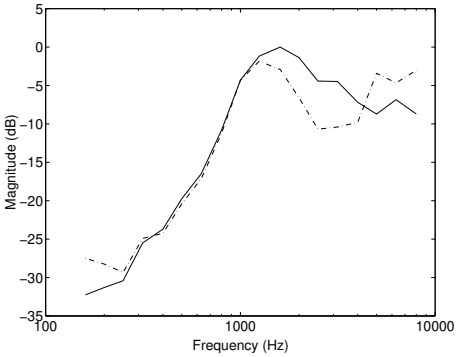


Figure 4.29: Frequency response of the hardware directional microphone (−) and the software directional microphone (−·).

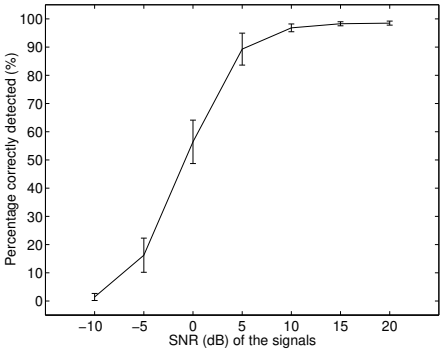


Figure 4.30: Average performance of the VAD for the speech-and-noise periods of all the signals (omnidirectional microphone, directional microphones (software and hardware), two-stage adaptive beamformer and optimal filtering technique) as a function of the SNR in stationary and non-stationary conditions.

quency response of the software directional microphone. Having more high frequency components, the signal of the hardware directional microphone has more temporal changes than the signal of the software directional microphone. This can explain the difference in the VAD-output between both directional microphones.

Similar to the results under stationary conditions, the best performance was revealed by connecting the VAD to the output of a noise reduction strategy. For a SNR $\geq$ -5dB, the speech-and-noise periods were well-detected. $Per > 90\%$ was observed for both microphone configurations when the VAD used the output signals of the two-stage adaptive beamformer and the optimal filtering technique. It seems that the frequency responses of the two directional microphones also have an influence on the VAD-output. Indeed, with the hearing aid DO, the VAD was very sensitive to the temporal variations of the multitalker babble and tended to detect only the speech-and-noise periods ($Per > 80\%$) whereas the noise-only periods were practically not detected, $Per \approx 30\%$ at a SNR of -10dB. With the hearing aid OO, Per was above 80% for the detection of the noise-only periods and Per was above 65% for the detection of the speech-and-noise periods. This means that in order to have a similar performance of the VAD, the sensitivity of the VAD algorithm needs to be fitted for every signal.

4.6.4.3 Influence of the SNR

The average performance of the VAD for the speech-and-noise periods as a function of the SNR of all signals (omnidirectional microphones, directional microphones and output of the noise reduction strategies) is illustrated in **figure 4.30**. In both stationary and non-stationary conditions, the performance of the VAD dropped for a SNR below 5dB.

The performance of the VAD in a multiple jammer sound scenario was also evaluated. A similar VAD-output was obtained for the different signals. Similar results for stationary conditions were obtained with the multitalker babble because the sum of the non-stationary signals at the omnidirectional microphones resulted in a stationary-like signal.

The VAD works on the log-energy of the signal, hence, its performance depends on the SNR. The better the SNR, the better the discrimination between the speech-and-noise period and the noise-only period. Connecting the VAD to the output of the noise reduction algorithm revealed the best performance. It corresponds well to the idea introduced by Lee and Ahn (2001). Indeed, Lee added a preprocessing step to improve the performance of the VAD. In our case, the preprocessing was the noise reduction algorithm itself. In the sequel, all the experiments will be performed with the VAD connected to the output of the noise reduction strategy.

4.7 Perceptual evaluations

This section presents the first perceptual evaluation of the SVD-based optimal filtering technique. The main perceptual research questions of this study are related to the optimization of the speech distortion parameter μ of the SVD-based optimal filtering technique (section 4.5.1), and, a more realistic comparison between the latter technique and the two-stage adaptive beamformer. The influence of the jammer sound scenes, the spectro-temporal character of the jammer sound and the microphone configuration of the hearing aids on the gain in speech intelligibility for the optimal filtering technique and the two-stage adaptive beamformer is investigated. The publications related to this section are (Maj et al., 2003c; Maj et al., 2004).

4.7.1 Method

The speech distortion parameter μ of the SVD-based technique was optimized for one jammer sound scene (speech at 0^o and noise at 90^o), one spectro-temporal character of the jammer sound (stationary speech weighted noise of the sentences) and one basic microphone configuration in the hear-

ing aid (DO). The first group of subjects (ten persons) performed twice the SRT-measurements for the omnidirectional microphone and the optimal filtering technique for different values of μ. The SRT-improvements of the noise reduction strategies were relative to the omnidirectional microphone. For the comparison of the two noise reduction algorithms in a single and in a multiple jammer sound scenario, the tests were performed for different spectro-temporal characters of the jammer sounds, and two basic microphone configurations in the hearing aid. SRT-measurements were carried out with two groups of subject consisting of ten and five normal hearing persons. The group of ten subjects (the same as for the optimization of the speech distortion parameter μ) performed the evaluation with the hearing aid DO and the group of five persons performed the evaluation with the hearing aid OO. The tests of the omnidirectional microphone, the two-stage adaptive beamformer and the SVD-based optimal filtering technique of hearing aids OO and DO were carried out in two spectro-temporally different jammer sounds (SW: unmodulated speech weighted noise and MB: multitalker babble) and in two different jammer sound scenarios (speech source at 0^o and a single noise at 90^o, versus speech source at 45^o and three uncorrelated noise sources at $90^o/180^o/270^o$). All the tests were carried out with the same speech material (sentences from a male speaker).

4.7.1.1 Experimental set-up

For the perceptual experiments, the same test room as described in the section 3.4.1 and the same hearing aids as presented in the section 4.6.3 were used.

The experimental set-up is illustrated **figure 4.31**. The BTE-devices were positioned on a mannequin in the test room and subjects evaluated the processed speech in the control room. Five loudspeakers (Yamaha CBX-S3) were used for the speech and discrete noise sources: one for the speech signal and three for the noise signals. The loudspeaker for the speech signal was situated in front of the mannequin at 0^o or at 45^o (side of the hearing aid) depending on the test scenario. The loudspeakers for the noise signals were located at 90^o (side of the hearing aid), at 270^o (the opposite side of the hearing aid) and behind the mannequin at 180^o. The loudspeakers were calibrated separately to obtain the same sound level for each loudspeaker in the middle of the five loudspeaker positions. The center of the loudspeakers was at the same height as the hearing aid of the mannequin (140cm) and the distance between each loudspeaker and the center of the mannequin's head was one meter. For noise scenarios with more than one spectrally identical jammer noise, the presented sources were uncorrelated.

In the control room, the speech signals were output by a computer sound card (Sound Blaster 16 bit) and the noise signals by a SONY CD991 CD player. For amplification purposes and intensity control, the speech signal

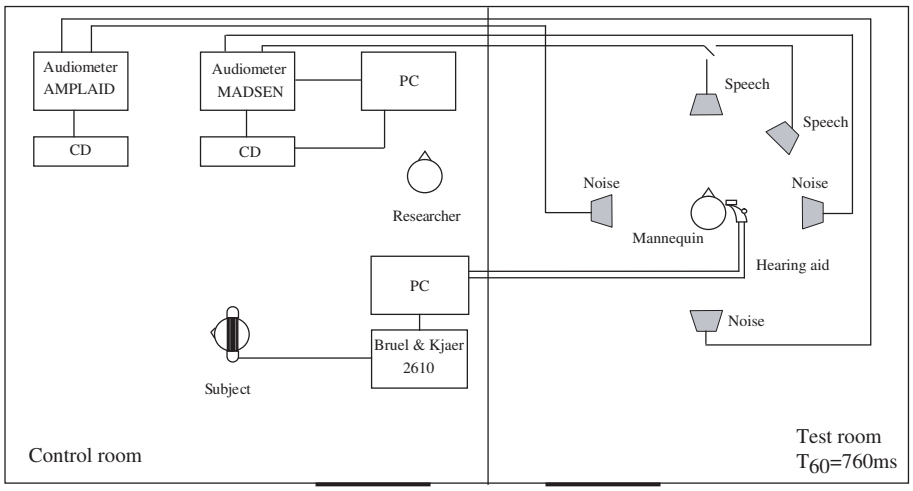


Figure 4.31: The experimental set-up for the perceptual evaluations of the SVD-based optimal filtering technique.

and the noise at 90^o were sent through a MADSEN OB822 audiometer, and the noises at 180^o and 270^o through an AMPLAID 309 audiometer. The signals of the microphones of the hearing aids were amplified with a two-channel Larson Davis 2200C amplifier and were digitized, at a sampling frequency of 12kHz using a LynxONE PC-sound card having two input channels with 16-bit analog-to-digital conversion (ADC). For the perceptual evaluation, C-code was developed to run the algorithms in real time. The output signal of the noise reduction algorithms was sent to an analog output of the LynxOne card. This analog signal was connected to a Bruel & Kjaer (2610) amplifier and presented monaurally to the subjects through headphones TDH39. This Bruel & Kjaer amplifier was calibrated to obtain an overall level of 65dBSPL on the right channel of the headphone when a level of 65dBPSL was present at the center of the loudspeakers with the head absent.

4.7.1.2 Subjects

To avoid eventual fatigue and/or training effects in the data, two different groups of normal hearing listeners participated in this study. The first group consisted of ten normal hearing listeners and their ages ranged from twenty to twenty seven (mean of twenty two years). This group performed the tests to optimize the speech distortion parameter μ of the SVD-optimal filtering technique and the tests with the hearing aid *DO* to evaluate the performance of the two noise reduction techniques. The second group consisted of five normal hearing persons and their ages varied from twenty to forty two (mean

of thirty two years). This group performed the tests with the hearing aid *OO* to evaluate the two noise reduction techniques. In both groups, the pure tone threshold was not less than or equal to 15dBHL at the octave frequencies from 125Hz to 8kHz.

4.7.1.3 Test materials

The adaptive procedure was used to measure the SRT and the level step size for each speech material was taken as 2dB. The estimated SRT was the average level of the last six responses of a list of thirteen sentences. The three noise sources were calibrated separately, to obtain a constant sound level at the reference point. Thus, a sound level of 65dBSPL and 69.8dBSPL was obtained with one noise source and three noise sources, respectively. Sentences spoken by a male voice (Versfeld et al., 2000) were used as speech materials, and, unmodulated speech weighted noise (with spectrum identical to the corresponding speech material) and multitalker babble were used as noise materials. One of the two test noises was unmodulated noise speech weighted according to the spectrum of the specific speech materials used. The other noise was multitalker babble, taken from the compact disk Auditory Tests (Revised) edited by Auditec of St. Louis.

4.7.2 Results

In this subsection, the results for the optimization of the speech distortion parameter μ of the SVD-based optimal filtering technique and the gain in speech intelligibility of both noise reduction algorithms for the two jammer sound source scenarios are presented.

4.7.2.1 Speech distortion parameter

Figure 4.32 shows the influence of the speech distortion parameter μ on the SRT-improvements of the SVD-based optimal filtering strategy. The best improvement of the speech intelligibility was obtained for $\mu = 1.75$. For $1 \leq \mu \leq 2$, the obtained SRT-improvements were similar (within 1dB), but the SRT-improvement decreased quite rapidly for $\mu = 4$ and $\mu = 10$. For these two values of μ, too much distortion was brought to the speech signal and affected the speech intelligibility. Perceptual measurements were also performed for $0 < \mu < 1$ and similar SRT-improvements, as obtained with $1 \leq \mu \leq 2$, were measured. Hence, the speech distortion parameter μ was fixed to 1.75.

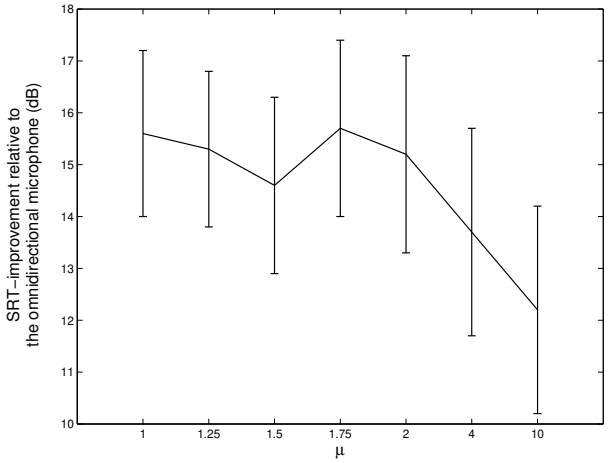


Figure 4.32: Influence of the speech distortion parameter μ on the SRT-improvements of the SVD-based optimal filtering technique.

4.7.2.2 Noise scenarios

Table 4.2 shows the improvements (in dB) of the SRT relative to the omnidirectional microphone for the two-stage adaptive beamformer and for the SVD-based optimal filtering technique and for both types of hearing aids (OO and DO). Every data cell corresponds to the mean (and the standard deviation) of all subjects in one group in each test condition. Columns three to four show the results of the SRT-improvement between the omnidirectional microphone and the two-stage adaptive beamformer, and column five to six show the results of the SRT-improvement between the omnidirectional microphone and the optimal filtering technique. This is presented for the two noise materials (SW and MB) and the two jammer sound scenes ($S_{0^\circ}N_{90^\circ}$ and $S_{45^\circ}N_{90^\circ/180^\circ/270^\circ}$). Statistic analysis were performed on the data using SPSS 10.0 software where general aspects and more detailed aspects were studied with repeated measures ANOVA and paired comparisons analysis, respectively.

4.7.2.2.1 Single noise source scenario

The analysis of within-subject contrasts showed that the SRT-improvements were significantly different between the noise reduction strategies ($F=26.70$, $p<0.001$) but no significant differences between the different spectro-temporal characters of the jammer sound ($F=0.43$, $p=0.525$). There was no interaction term between the algorithms and the spectro-temporal character of the jammer sound ($F=12.08$, $p=0.004$), and no interaction terms between the microphone configuration in the hearing aids and the noise materials ($F=0.08$,

Table 4.2: The differences in SRT (dB) averaged (mean(STD)) per group of subjects between the two-stage adaptive beamformer and the SVD-based optimal filtering technique, respectively, and the omnidirectional microphone for different conditions. Two jammer source sound scenarios were studied. A first with a speech signal at 0^o and a single noise source at 90^o (side of the hearing aid). A second, with a speech signal at 45^o and three uncorrelated noises at $90^o/180^o/270^o$. A group of ten normal hearing listeners were tested with the hearing aid *DO* and a second group of five normal hearing listeners were tested with the hearing aid *OO*.

		Adaptive beamformer		Optimal filtering technique	
Noise scenario	Noise material	DO	OO	DO	OO
S_{0° N_{90°	SW	15.8 (1.8)	14.3 (2.2)	15.1 (1.3)	14.0 (2.4)
	MB	17.0 (2.7)	14.3 (1.3)	13.3 (3.0)	12.6 (1.7)
S_{45° $N_{90^\circ/180^\circ/270^\circ}$	SW	7.5 (0.9)	6.5 (1.8)	9.0 (1.5)	8.4 (1.5)
	MB	5.8 (2.8)	4.7 (2.4)	6.3 (2.1)	5.2 (1.9)

p=0.784) or the noise reduction techniques (F=4.39, p=0.056).

• *Hearing aid DO*: Important SRT-improvements were obtained: 15.8dB and 15.1dB in stationary conditions, and, 17.0dB and 13.3dB in non-stationary conditions, for the adaptive beamformer and the optimal filtering technique respectively. The behaviour of both noise reduction strategies differed in multitalker babble. The paired comparisons analysis gave a significant difference between the adaptive beamformer and the optimal filtering technique (p<0.001). In non-stationary conditions the difference in SRT-improvement between the two strategies was 3.7dB.

• *Hearing aid OO*: With the adaptive beamformer, a similar SRT-improvement of 14.3dB was obtained in stationary and non-stationary conditions. With optimal filtering, the SRT-improvements were 14.0dB and 12.6dB in speech weighted noise and multitalker babble, respectively. There were no significant differences between these SRT-improvements (p=0.370). Unlike for the hearing aid *DO*, there were no significant differences between the SRT-improvement obtained with the two-stage adaptive beamformer and the SVD-based technique in multitalker babble (p=0.088). Paired comparisons were carried out and showed that the performance of the two-stage adaptive beamformer was dependent on the microphone configuration in the hearing aids (p=0.023), but that this was not the case for the optimal filtering technique (p=0.399).

4.7.2.2.2 Multiple noise sources scenario

The analysis of within-subject contrasts showed the SRT-improvements were significantly different between the two spectro-temporal characteristics of the jammer sound (F=10.62, p=0.006) and the two noise reduction strategies (F=13.79, p=0.003). There were no interaction terms between the algorithms and the spectro-temporal character of the jammer sound (F=3.41, p=0.087), between the microphone configuration in the hearing aids and the noise materials (F=0.09, p=0.775) and between the two microphone configurations in the hearing aids and the noise reduction techniques (F=0.10, p=0.752).

• *Hearing aid DO*: The SVD-based optimal filtering technique was significantly better than the two-stage adaptive beamformer when a stationary speech weighted noise was present (p=0.005). SRT-improvements of 7.5dB and 9.0dB were obtained with the adaptive beamformer and the optimal filtering technique, respectively. In multitalker babble, the speech intelligibility improvement of both strategies was similar (p=0.452): SRT-improvement of 5.8dB was obtained for the adaptive beamformer and 6.3dB for the optimal filtering technique.

• *Hearing aid OO*: Similar results were obtained with the device *OO*. There was a significant difference between the optimal filtering technique and the adaptive beamformer in stationary conditions (p=0.011) but not in non-stationary conditions (p=0.532). The SRT-improvements were 6.5dB and 4.7dB with the two-stage adaptive beamformer, and, 8.4dB and 5.2dB with the SVD-based optimal filtering technique, in speech weighted noise and multitalker babble, respectively. A paired comparison showed that the two different microphone configuration did not have an influence on the performance of the two-stage adaptive beamformer (p=0.212) and the optimal filtering technique (p=0.187).

4.7.2.3 Discussion

4.7.2.3.1 Single noise scenario

The single noise scenario was optimal for the two-stage adaptive beamformer because the desired target was in the look direction given to the beamformer. Although a leakage of the speech signal into the noise reference of the adaptive beamformer occurred (the first filter was calibrated in anechoic conditions and fell short in reverberant conditions), the presence of the speech signal in the noise reference was limited (**figure 3.13**). The tests were carried out at average SNRs between -10dB and 0dB (at the omnidirectional microphone) in stationary and non-stationary conditions, respectively, and the detection performance of the VAD for the speech-and-noise periods was high enough to avoid that the ANC stage was corrupted by the statistics of the speech signal. Hence, the two-stage adaptive beamformer gave important SRT-improvements in stationary conditions, 15.8dB and 14.3dB, or 17.0db and 14.3dB in non-stationary

conditions, with the hearing aid DO and OO respectively.

With the optimal filtering technique, the VAD discriminated accurately the speech-and-noise periods ($Per > 75\%$) from the noise-only periods ($Per > 90\%$) in stationary conditions for both microphone configuration and in non-stationary conditions with the hearing aid OO. The perceptual tests were performed at a SNR of about -10dB in stationary condition and about -5dB in non-stationary conditions. The SRT-improvements were 15.1dB and 14.0dB in stationary conditions with the hearing aids DO and OO, respectively, and 12.6dB in non-stationary conditions with the hearing aid OO. In these noise scenarios, no significant differences were obtained between the two noise reduction strategies. With the hearing aid DO in the multitalker babble a significant difference was found between the two noise reduction strategies. In this configuration, the VAD was sensitive to the temporal variation of the noise and tended to detect only speech-and-noise periods even when the noise-only periods were present. At a SNR of -10dB, similar to the SNR to which the perceptual tests were performed, the VAD detected correctly about 95% the speech-and-noise periods and only 25% of the noise-only periods (**figure 4.28**). On the one hand, this did not result in a cancellation of the desired signal but on the other hand the optimal filtering technique did not have a good estimate of the noise statistics and hence could not perform an optimal noise cancellation. The function based on the derivative D_F had a moderate influence on the performance of the optimal filtering technique because the speech-and-noise periods were well detected by the VAD and the threshold T_D was fitted low enough so that the spectral changes of the multitalker babble did not make the derivative D_F effective ($D_F > T_D$).

4.7.2.3.2 Multiple noise sources scenario

In stationary conditions, the optimal filtering technique gave a better SRT-improvement than the two-stage adaptive beamformer. The differences were small between the two strategies (1.5dB and 1.9dB with the hearing aids DO and OO, respectively) but very important for hearing-aid users. An improvement of 1dB in SNR can concur to an increase of speech understanding of about 15% in every day speech communication (Plomp and Mimpen, 1979). The noise scenario was not optimal for the beamformer technique because the desired signal was not in the look direction. The amount of speech signal leakage into the noise reference increased (**figure 4.6** and **4.17**) and hence the two-stage adaptive beamformer became more sensitive to the errors of the VAD (**figures 4.7** and **4.15**). Unlike the adaptive beamformer, the performance of the optimal filtering technique is not a function of the position of the speaker because this technique works without assumptions on the desired target.

In non-stationary conditions, the optimal filtering technique performed as well as the two-stage adaptive beamformer. The temporal changes of the signal were to some extent attenuated by summing the three noise sources. In

this configuration, the VAD-output was similar to the output under stationary conditions, but significant differences were found between the two spectro-temporal characters of the noise. The asymptotic maximum noise reduction performance of the noise reduction strategies is reached if the time constants of the techniques are small enough to track the temporal and spatial changes of the noises during the tests. Under stationary conditions, the spectral and the spatial variations are minimal and in non-stationary conditions with a single noise source only the spectral changes need to be tracked. The time constant was small enough ($<$ 1s) to follow the changes in these noise scenarios. In a multiple noise scenario with non-stationary noise, the time constant of the noise reduction strategies was certainly not small enough to track the variations in the spectral and the spatial domains. The technique could not perform optimal noise reduction and the asymptotic maximum noise reduction performance was not reached.

Relative to the single noise source scenario, the SRT-improvements of the noise reduction strategies decreased by between 6 and 10dB. Peterson (1989) analyzed the effects of the number of noise sources and microphones on the performance of the noise reduction algorithms: the more independent sensor inputs the better noise sources can be nulled out with appropriate signal processing. Thus, two microphone configurations are only optimal for the cancellation of just one noise source.

4.7.2.3.3 Microphone configuration

With the two-stage adaptive beamformer, the microphone configuration of the hearing aid had an influence on the performance of the strategy. The hearing aid *DO* gave additional improvement relative to the hearing aid *OO*. These results agree with a previous study, but a larger difference between both types of microphone configurations was observed (Maj et al., 2003). The tests were carried out in the same conditions as in this study (same speech material, speech weighted noise, noise scenario and room configuration), and a difference of 3.6dB was obtained between the devices *OO* and *DO* for the adaptive beamformer. In this study, the difference was only 1.5dB between both devices. This dissimilarity in the results may come from the positions of the microphones which were different in the hearing aid for both studies. The performance of the SVD-based optimal filtering technique was not dependent on the microphone configuration in the hearing aid.

4.7.2.4 Computational complexity

The computational complexity of the two-stage adaptive beamformer equals $(2.N_1 + 7.N_2) \times fs$ operations (multiplications or additions) per second (ops/s) where N_1 represents the size of the fixed filter, N_2 the size of the adaptive filter and fs the sampling frequency. In this study, the two-stage

adaptive beamformer had a computational complexity of 2Mops/s, with $N_1 = 8$ taps and $N_2 = 22$ taps and a sampling frequency of 12kHz. The computational cost of the SVD-based optimal filtering equals $(17.5.N^2+4.N^2) \times fs$ ops/s. Here N represents the size of the filters per channel and was set to $N = 15$. Therefore the computational cost equals 232Mops/s at 12kHz. SVD procedures have a high computational complexity, but, recent studies showed that this complexity can be further reduced, making this approach attractive for practical systems. A subband approach or alternative algorithms based on a QR Decomposition were proposed in (Rombouts, 2003) and (Spriet et al., 2002). More recently, a LMS approach was proposed and had approximately the same computational cost as the two-stage adaptive beamformer, be it at the expense of a larger memory requirement (Spriet et al., 2003b; Spriet et al., 2003c).

4.8 Conclusion

In this chapter, an optimal filtering technique was developed for dual microphone hearing aid applications. The technique was computed through a SVD-based method and the study demonstrated that an additional gain in speech intelligibility is obtained in complicated noise scenarios relative to the two-stage adaptive beamformer. The optimal filtering technique was assessed theoretically, physically and perceptually.

The *theoretical evaluation* demonstrated that the optimal filtering technique seems to be more robust than the two-stage adaptive beamformer against gain variations of the microphones in hearing aids. Both strategies are sensitive to the errors of the VAD. The performance of the two-stage adaptive beamformer depends on the position of the speaker. When the speaker is not in the look direction of the algorithm, leakage of the speech signal into the noise reference occurs. Consequently, the desired signal is partially cancelled at the output of the strategy. The performance of both noise reduction techniques decreases when the number of jammer sound sources increases.

The *physical evaluation* was well predicted by the theoretical evaluation. Indeed, both experiments demonstrated that the errors of the VAD have an important influence on the performance of the noise reduction strategies. When the noise periods are wrongly detected as speech-and-noise periods, the statistic of the noise signal are corrupted by the statistics of the speech signal and a cancellation of the desired signal is observed. A method to improve the performance of the VAD was proposed. By connecting the voice activity detection algorithm at the output of the noise reduction technique, a good discrimination between the speech-and-noise periods and the noise-only periods even at low SNRs was achieved. The experiments revealed that the optimal filtering technique had roughly the same behaviour as the two-stage adaptive

beamformer in different single or multiple jammer sound source scenarios but seems to give a better gain in speech intelligibility through the G_{AI} in a diffuse noise scenario.

For *the perceptual experiments*, both noise reduction strategies were implemented in real time on a PC-platform. A technique based on a Jacobi-type GSVD-updating algorithm was used to approximate the optimal filtering technique. A constraint was applied on the optimal filtering technique to avoid temporal distortions of the output signal when VAD errors occur. This constraint prevented that the distance between the generalized singular values became too small which otherwise results in a zeroing of the coefficients of the filters. The perceptual experiments were carried out in two different noise scenarios with two microphone configurations. The SVD-based optimal filtering technique was compared to the two-stage adaptive beamformer. The perceptual experiments demonstrated that, in an optimal noise scenario for the two-stage adaptive beamformer, the optimal filtering technique performs as well as the former technique. Significant SRT-improvements were obtained for both noise reduction strategies because single jammer sound source scenarios are appropriate for dual microphone hearing aids. In a multiple noise scenario, the optimal filtering technique performed better than the adaptive beamformer in stationary conditions. The two-stage adaptive beamformer outperforms the noise reduction algorithms implemented in modern digital hearing aids (chapter 1) and this study demonstrated that the optimal filtering technique outperforms the adaptive beamformer in adverse listening conditions.

Chapter 5

Conclusion

Hearing-impaired listeners and hearing aid users have great difficulty understanding speech in a noisy environment. This problem has motivated the development and the use of noise reduction algorithms in hearing aids. In this dissertation, two noise reduction algorithms were developed for dual microphone hearing aids and assessed theoretically, physically and perceptually. The first noise reduction strategy was based on a GSC procedure and the second was based on a multimicrophone optimal filtering technique. In this chapter, the main conclusions are summarized and suggestions for future research and experiments are given.

5.1 Two-stage adaptive beamformer

The aim of this study was to optimize an adaptive beamformer, namely a two-stage adaptive beamformer technique, for a real time implementation in digital hearing aids.

For robustness, the first filter of the two-stage adaptive beamformer was kept fixed for a specific direction (0^o) and was calibrated for each and every hearing aid. Thus, the difference in the characteristics of the microphones was taken into account to minimize the presence of the speech signal in the noise reference of the strategy. The second filter was adaptive and attempted to model the noise signal in the speech reference during noise-only periods and subtracted noise from speech-plus-noise when speech-and-noise periods were present. A VAD was used to decide whether a speech-and-noise period or a noise-only period was present.

The theoretical evaluation demonstrated that the distance between the microphones and the internal noise of the sensors have an important influence on the performance of the adaptive beamformer. By increasing the distance between the microphones and by decreasing the internal noise of the sensors, the adaptive beamformer performs better. However, the space between the sensors is limited by the housing of the BTE hearing aids and a short distance between two omnidirectional microphones may affect the sensitivity of the software directional microphone at the low frequencies. The internal noise can be reduced by taking care of the characteristics of the microphones inside the hearing aids. The theoretical evaluation showed that the speech signal is not distorted by the processing of the two-stage adaptive beamformer when the speaker is in the look direction of the beamformer.

The two-stage adaptive beamformer technique was optimized for a real time implementation in a wearable research platform. This platform had computational power similar to that available in modern digital hearing aids. The size of the filters was optimized by minimizing the presence of the speech signal in the noise reference of the adaptive beamformer and by maximizing the gain in intelligibility weighted SNR (SNR_{SII}) at the output of the strategy relative to an omnidirectional microphone. A hearing aid with a hardware directional microphone and an omnidirectional microphone was considered in this section and a filter length of ten and thirty at a sampling frequency of 15.625kHz was chosen for the first and the second stage, respectively. A physical evaluation was carried out in a reverberant room ($T_{60} \approx 450ms$) and showed that a significant gain in SNR_{SII} could be obtained for different locations of a noise source. Also, the evaluation confirmed that the two-stage adaptive beamformer does not distort the speech signal during processing when the speaker is at 0^o and with a perfect VAD. The real time implementation of the adaptive beamformer was validated by carrying out a perceptual evaluation and a significant SRT-improvement of about 11dB was obtained by the adaptive beamformer relative to a hardware directional microphone. The SRT measurements were carried out in a moderately reverberant room ($T_{60} \approx 350ms$) with a speech source at 0^o (in front of the mannequin) and a noise source at 90^o (on the side of the hearing aid). The hearing aid was worn by a mannequin and the output signal of the directional microphone and the two-stage adaptive beamformer were presented monaurally to the subjects through headphones. The SRT-improvement measured in the perceptual evaluation agreed with the gain in SNR_{SII} calculated in the physical evaluation.

A large scale evaluation was performed in moderately reverberant environments which are similar to real living rooms. Two microphone configurations in hearing aids were used. Seventeen normal hearing listeners and five hearing impaired subjects participated in this study. The hearing aids were worn by the subjects and the SRT measurements were performed for the omnidirectional microphones, the directional microphones (software or hardware)

and the two-stage adaptive beamformer in different noise scenarios. The SRT-improvement of the directional microphones or the noise reduction strategy were relative to the omnidirectional microphones. On the one hand, the experiments demonstrated that the SRT-improvements did not depend on the spectro-temporal character of the jammer sound as well as the speech sounds, although the absolute value of the SRT widely differed. On the other hand, the SRT-improvements depended on the jammer sound source scenario. The higher the complexity of the jammer sound scene, the lower the SRT-improvements. In a simple jammer sound source scenario, the adaptive beamformer brought a clear benefit in speech intelligibility relative to the directional microphones. However, in diffuse noise, the performance of the adaptive beamformer falls back to the effect of the directional microphone. This is why a noise reduction strategy in hearing aid applications, should have fixed processing to be effective in high reverberant conditions and adaptive processing to be optimal in low reverberant conditions. The experiments also demonstrated that the same advantage was obtained for hearing impaired with the software directional microphone and the adaptive beamformer. These results need to be considered with some caution because of the small number of hearing impaired subjects. Finally, the experiments revealed that the microphone configuration has an impact on the performance of the adaptive beamformer. A clear benefit was obtained by the hearing aid using a hardware directional microphone and an omnidirectional microphone over the hearing aid using two omnidirectional microphones. The perceptual results were in fair accordance with the physical evaluations. This relatively good agreement of physical and perceptual measures (within approximately 1dB) in this study showed that a very good prediction of the SRT-improvement could be obtained using physical measurements in realistic conditions (on a mannequin, in realistic reverberation and SII calculation).

Finally, the two-stage adaptive beamformer was implemented in a commercial device and compared to an adaptive directional microphone, similar to the state-of-the-art noise reduction in modern commercial digital hearing aids. The characteristics of the microphones vary from sensor to sensor and have a direct impact on the performance of the fixed processing systems. Physical experiments demonstrated the importance of an individualized calibration procedure for fixed filter (e.g. software directional microphone) to obtain a similar noise reduction performance for every hearing aid using the same microphone configuration. A perceptual experiment demonstrated that the two-stage adaptive beamformer always performed better than the adaptive directional microphone and a software directional microphone in single jammer sound source scenarios, whereas in multiple noise source scenario the adaptive strategies (the directional microphone and the beamformer) performed as well as the fixed directional microphone.

5.2 SVD-based optimal filtering technique

In this study, a multimicrophone optimal filtering technique was developed for dual microphone hearing aid applications. The optimal filtering technique works without a priori assumption about the microphone signals and the speaker position, unlike the two-stage adaptive beamformer. The comparison between both techniques was carried out through a theoretical, physical and perceptual evaluation.

The theoretical experiments demonstrated that the performance of the optimal filtering technique and the two-stage adaptive beamformer is a function of the characteristics of the microphones (gain and phase) and the VAD errors. The optimal filtering technique seemed to be less sensitive to the variations of the characteristic of the microphones but necessitated a robust VAD to enhance the effectiveness of the strategy. In multiple noise source scenarios, both noise reduction strategies had roughly the same behaviour when a perfect VAD was considered.

The optimal filter was computed by means of a SVD-based technique. The different parameters of the proposed strategy were optimized in the physical evaluation by maximizing the gain in intelligibility weighted SNR at the output of the strategy relative to an omnidirectional microphone. The SVD-based optimal filtering technique was compared to the adaptive beamformer in different jammer sound source scenarios. The physical experiments revealed that both noise reduction strategies have the same behaviour in a single or a multiple jammer sound source scenario when a perfect VAD was considered. Also, the experiments demonstrated that the SVD-based optimal filtering technique does not distort the speech signal during processing. In a diffuse noise case, the optimal filtering technique was seen to outperform the adaptive beamformer. In the presence of VAD errors, the performance of both strategies dropped significantly, especially when the statistics of the noise signal were corrupted by the statistics of the speech signal. On the one hand the performance of the adaptive beamformer technique was dependent on the speaker position. The more the direction of the speaker differs from 0^o, the higher the level of the speech signal in the noise reference, resulting in a more severe cancellation of the desired signal at the output of the strategy. On the other hand, the performance of the optimal filtering technique was independent of the speaker position.

A method was proposed to improve the performance of the VAD. The VAD was connected at the signal with the highest SNR, namely the output of the noise reduction strategy. The physical evaluation showed that a good discrimination between the speech-and-noise periods and the noise-only periods of the noisy speech signals was performed even for low SNRs.

For the perceptual experiments, both noise reduction strategies were implemented in real time on a PC-platform. A recursive algorithm based on a Jacobi-type GSVD-updating was used to approximate the optimal filtering technique. When VAD errors occur, the SVD-based optimal filtering technique may introduce temporal distortions to the output signal. By detecting noise-only periods instead of speech-and-noise periods, the generalized singular values converge to the same values, resulting in the cancellation of the output signal. To avoid temporal distortions, a constraint on the VAD was applied as a function of a time derivative D_F, which is based on a distance measure between the generalized singular values of the matrices $\mathbf{U}_k$ and $\mathbf{N}_k$. Depending on the value of D_F, the decision of the VAD was kept valid or was modified.

The perceptual experiments demonstrated that the SVD-based optimal filtering technique could perform as well as the two-stage adaptive beamformer in a single noise scenario, i.e. the ideal scenario for the former technique, and outperform the adaptive beamformer in a multiple noise scenario. In the single noise source scenario, the speech source was in the look direction of the beamformer, and this limited the presence of the speech signal in the noise reference of the two-stage adaptive beamformer. An important gain in speech intelligibility was obtained in stationary and in non stationary conditions with the latter strategy. The performance of the optimal filtering technique depended on the VAD. When good discrimination between the speech-and-noise periods and the noise-only periods was achieved, the optimal filtering technique performed as well as the two-stage adaptive beamformer. When the VAD was too sensitive to the temporal variation and tended to detect only speech-and-noise periods in non-stationary conditions, the noise reduction worked with a bad estimate of the statistics of the noise signal and could not achieve an optimal noise cancellation. In the multiple jammer sound source scenario, and with the speaker not in the look direction of the adaptive beamformer, the level of the speech signal in the noise reference of the adaptive beamformer increased. Thus, the latter technique became more sensitive to the errors of the VAD and a cancellation of the speech signal could occur in the output of the strategy. In stationary conditions, the optimal filtering technique outperformed the two-stage adaptive beamformer and in non-stationary conditions a similar behaviour could be observed for both strategies. The time constant of the adaptive noise reduction algorithms was not small enough to track the spectral and spatial changes to achieve an optimal noise reduction.

5.3 Suggestions for further research

5.3.1 Two-stage adaptive beamformer

The performance of the two-stage adaptive beamformer depends mainly on the robustness of the first filter. The initialization procedure was a function of the microphone characteristics and the position of the speaker. Thus, the first filter was adapted in anechoic conditions to each and every hearing aid, which was not a simple task. When the assumptions (anechoic conditions, speaker position at 0^o, matching of the microphones) are violated, a leakage of the speech signal in the noise reference occurs, resulting in partial cancellation of the speech signal at the output of the noise reduction strategy.

Several strategies have been proposed to improve the robustness of the fixed filter, and this, without calibration procedures. Cox et al. (1987) introduced a norm-constraint on the adaptive filter of the beamformer to avoid the cancellation of the desired signal. Another method (Hoshuyama et al., 1999) consisted of avoiding the presence of the desired signal in the noise reference by applying adaptive filters between the output of the fixed beamformer of the GSC procedure and the noise reference. Recently, a procedure to design broadband beamformers taking into account the tolerances in the microphone positions was presented (Doclo, 2003). This method requires the distribution function of the microphone position errors. A complete analysis and comparisons of these methods may be the subject of future work.

5.3.2 Optimal filtering technique

The optimal filtering solution was computed by means of a SVD-based algorithm which led to a high computational complexity. Recent studies showed that the computational complexity can be reduced significantly by using a stochastic-gradient approach, which has approximately the same computational cost of calculation as the two-stage adaptive beamformer (Spriet et al., 2003b; Spriet et al., 2003c). As a future work, the latter method should be implemented in real time and a comparison between the SVD and the stochastic-gradient approaches with perceptual measurements should be performed. It is to be expected that the stochastic-gradient procedure will perform as well as the SVD procedure, but for a significantly reduction in cost of calculation.

The optimal filtering technique seems to be less sensitive to the variations of the characteristics of the microphones than the two-stage adaptive beamformer. This was demonstrated theoretically and physically (section 4.2) (Spriet et al., 2001; Doclo, 2003) but not perceptually. As a future work, perceptual evaluations should be performed to prove that the optimal filtering technique presents a more robust solution than the adaptive beamformer against the different characteristics of the microphones.

In this dissertation, the strategies developed relied on a VAD to discriminate the speech-and-noise periods from noise-only periods. However, the VAD errors were seen to have an important impact on the noise reduction performance, especially at low SNR and in non-stationary conditions. The performance of the VAD used in this study might be improved by combining different methods of VAD (zero crossing rate, periodicity estimator method, Log-likelihood) and by integrating a distance measure on the speech-and-noise statistics and the noise-only statistics such as the derivative D_F introduced in this study.

Bibliography

Affes, S. and Grenier, Y. (1997). A signal subspace tracking algorithm for microphone array processing of speech. *IEEE Transactions on Speech and Audio Processing*, ASSP-5(5):425–437.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control*, AC-19:716–723.

Alcantara, J., Moore, B., Kuhnel, V., and Launer, S. (2003). Evaluation of the noise reduction system in a commercial digital hearing aid. *International Journal of Audiology*, 42(1):34–42.

Anemuller, J. and Kollmeier, B. (2003). Adaptive separation of acoustic sources for anechoic conditions: A constrained frequency domain approach. *Speech Communication*, 39:79–95.

ANSI (1969). Methods for the calculation of the articulation index. *Acoustical Society of America*, American National Standard:ANSI S3.5–1969.

ANSI (1997). Methods for calculation of the speech intelligibility index. *Acoustical Society of America*, American National Standard:ANSI S3.3–1997.

Bachler, H. and Vonlanthen, A. (1995). Traitement du signal audio-zoom pour améliorer la communication dans le bruit. *Phonak Focus*, 18.

Baer, T., Moore, B. C., and Kluk, K. (2002). Effects of low pass filtering on the intelligibility of speech in noise for people with and without dead regions at high frequencies. *Journal of the Acoustical Society of America*, 112(3):1133–1144.

Beranek, L. L. (1954). *Acoustics*. New York: McGraw Hill Electrical and Electronic Engineering Series.

Beritelli, F., Casale, S., and Cavallaro, A. (1999). A multi-channel speech/silence detector based on time delay estimation and fuzzy classification. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 15-19, Phoenix, USA:93–96.

Bitzer, J., Simmer, K. U., and Kammeyer, K. D. (2001). Multi-microphone noise reduction techniques as front-end devices for speech recognition. *Speech communication*, 34:3–12.

Boll, S. F. (1979). Suppression of acoustic noise in speech using a spectral subtraction. *IEEE Transaction on Acoustics, Speech and Signal Processing*, ASSP-27(2):113–120.

Buck, M. (2002). Aspects of first-order differential microphone arrays in the presence of sensor imperfections. *European Transaction on Telecommunications, special issue on Acoustic Echo and Noise Control*, 13(2):115–122.

Buckley, K. M. (1986). Broad-band beamforming and the generalized sidelobe canceller. *IEEE Transaction on Acoustics, Speech and Signal Processing*, ASSP-34(5):1322–1323.

Bustamante, D. K., Worrall, T. L., and Williamson, M. J. (1989). Measurement and adaptive suppression of acoustic feedback in hearing aids. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 23-26, Glasgow, Scotland:2017–2020.

Byrne, D., Dillon, H., and Tran, K. (1994). An international comparison of long-term average speech spectra. *Journal of the Acoustical Society of America*, 96(4):2108–2120.

Caspary, O. and Maj, J. B. (1999). Détection de sinusoides noyées dans du bruit par les régions d'inclusion des valeurs propres. *Groupe de Recherche En Traitement du Signal et Image (GRETSI)*, September 13-17, Vannes, France:CD.

Cho, Y. D., Al-Naimi, K., and Kondoz, A. (2001). Improved voice activity detection based on a smoothed statistical likelihood ratio. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 7-11, Salt Lake City, USA:CD.

Clarkson, P. and Bahgat, S. (1991). Envelope expansion methods for speech enhancement. *Journal of the Acoustical Society of America*, 89(3):1378–1382.

Couch, L. W. (1997). *Digital and Analog Communication Systems*. Prentice Hall.

Cox, H., Zeskind, R. M., and Kooij, T. (1986). Practical supergain. *IEEE Transactions on Speech and Audio Processing*, ASSP-34(3):393–398.

Cox, H., Zeskind, R. M., and Owen, M. M. (1987). Robust adaptive beamforming. *IEEE Transactions on Speech and Audio Processing*, ASSP-35(10):1365–1376.

Cox, R. M. and Alexander, G. C. (1991). Hearing aid benefit in everyday environments. *Ear and Hearing*, 12(2):127–139.

Desloge, J. G., Rabinowitz, W. M., and Zurek, P. M. (1997). Microphone-array hearing aids with binaural output- part 1: Fixed-processing systems. *IEEE Transactions on Speech and Audio Processing*, SAP-5(6):529–542.

Doclo, S. and De Clippel, E. (1997). *Verbetering van spraaverstaan bij hoortoestellen via adaptieve ruisonderdrukking in reele tijd.* Internal report, Katholieke Universiteit Leuven.

Doclo, S. and Moonen, M. (2002). GSVD-based optimal filtering for single and multiple speech enhancement. *IEEE Transactions on Signal Processing*, SP-50(9):2230–2244.

Doclo, S. (2003). *Multi-microphone noise reduction and dereverberation techniques for speech applications.* PhD thesis, Katholieke Universiteit Leuven.

Douglas, S. C. and Sun, X. (2003). Convolutive blind separation of speech mixtures using the natural gradient. *Speech Communication*, 39:65–78.

Duquesnoy, A. J. and Plomp, R. (1983). The effect of a hearing aid on the speech-reception threshold of hearing-impaired listeners in quiet and in noise. *Journal of the Acoustical Society of America*, 73(6):2166–2173.

Ephraim, Y. and Malah, D. (1984). Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech and Signals Processing*, ASSP-32(6):1109–1121.

Ephraim, Y. and Van Trees, L. (1995). A signal subspace approach for speech enhancement. *IEEE Transactions on Acoustics, Speech and Signals Processing*, ASSP-3(4):251–266.

Figueiredo, J. C., Abel, S. M., and Papsin, B. C. (2001). The effect of the audallion beamformer noise reduction preprocessor on sound localization for cochlear implant users. *Ear and Hearing*, 22(6):539–547.

Friedmann, D. H. (1978). Pseudo-maximum-likelihood speech pitch extraction. *IEEE Transactions on Acoustics, Speech and Signals Processing*, ASSP-25(3):213–221.

Frost, O. L. (1972). An algorithm for linearly constrained adaptive array processing. *Proceedings of the IEEE*, 60:926–935.

Gannot, S. and Moonen, M. (2001). Subspace methods for multi-microphone speech dereverberation. *International Workshop on Acoustic Echo and Noise Control (IWAENC)*, September 10-13, Darmstadt, Germany:47–50.

Godara, L. C. and Gray, D. A. (1989). A structured gradient algorithm for adaptive beamforming. *Journal of the Acoustical Society of America*, 86(3):1040–1046.

Godara, L. C. (1991). Adaptive beamforming in the presence of correlated arrivals. *Journal of the Acoustical Society of America*, 89(4):1730–1736.

Golub, G. H. and Loan, C. F. V. (1996). *Matrix Computations*. MD: John Hopkins University Press, Baltimore, 3rd edition.

Greenberg, J. E. and Zurek, P. M. (1992). Evaluation of an adaptive beamforming method for hearing aids. *Journal of the Acoustical Society of America*, 91(3):1662–1676.

Greenberg, J. E., Peterson, P. M., and Zurek, P. M. (1993). Intelligibility-weighted measures of speech-to-interference ratio and speech system performance. *Journal of the Acoustical Society of America*, 94(5):3009–3010.

Greenberg, J. E. (1998). Modified LMS algorithms for speech processing with an adaptive noise canceller. *IEEE Transaction on Speech and Audio Processing*, SAP-6(4):338–351.

Greenberg, J. E., Desloge, J. G., and Zurek, P. M. (2003). Evaluation of array-processing algorithms for a headband hearing aid. *Journal of the Acoustical Society of America*, 113(3):1646–1657.

Griffiths, L. J. and Jim, C. W. (1982). An alternative approach to linearly constrained adaptive beamforming. *IEEE Transaction on Antennas Propagation*, AP-30:27–34.

Hansen, P. S. K. (1997). *Signal subspace methods for speech enhancement*. PhD thesis, Technical University Lyngby, Denmark.

Hansen, P. C. and Jensen, S. H. (1998). FIR filter representations of reduced-rank noise reduction. *IEEE Transaction on Signal Processing*, SP-46(6):1737–1741.

Hawkins, B. and Yacullo, W. (1984). Signal-to-noise ratio advantage of binaural hearing aids and directional microphones under different levels of reverberation. *Journal of Speech and Hearing Disorders*, 49(3):278–286.

Haykin, S. (1996). *Adaptive Filter Theory*. Prentice-Hall.

Heitkamper, P. (1997). An adaptation control for acoustic echo cancellers. *IEEE Signal Processing Letters*, 4:170–172.

Herault, J. and Jutten, C. (1986). Space or time adaptive signal processing by neural network models. *Neural Networks for Computing: AIP Conference Proceedings*.

Hoffman, M. W., Trine, T. D., Buckley, K. M., and Tasell, D. J. V. (1994). Robust adaptive microphone array processing for hearing aids: realistic speech enhancement. *Journal of the Acoustical Society of America*, 96(2):759–770.

Hoshuyama, O., Sugiyama, A., and Hirano, A. (1999). A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters. *IEEE Transaction on Signal Processing*, SP-47(10):2497–2506.

Huang, J. and Zhao, Y. (1997). Energy-constrained signal subspace method for speech enhancement and recognition. *IEEE Signal Processing Letters*, 4(10):283–285.

ICRA (1997). International collegium of reabilitative audiology. *Noise Signals ICRA*, (Ver 0.3):CD.

Jensen, S. H., Hansen, P. C., Hansen, S. D., and Sorensen, J. A. (1995). Reduction of broad band noise in speech by truncated QSVD. *IEEE on Transaction in Speech and Audio Processing*, SAP-3(6):439–448.

Joson, H. A. L., Asano, F., Suzuki, Y., and Stone, T. (1993). Adaptive feedback cancellation with frequency compression for hearing aids. *Journal of the Acoustical society of America*, 94(6):3248–3254.

Junqua, J. C., Reaves, B., and Mak, B. (1991). A study of endpoint detection algorithms in adverse conditions: incidence on a DTW and HMM recognize. *Eurospeech*, September 24-26, Genova, Italy:1371–1374.

Junqua, J. C., Mak, B., and Reaves, B. (1994). A robust algorithm for word boundary detection in the presence of noise. *IEEE on Transaction in Speech and Audio Processing*, SAP-2(3):406–412.

Kates, J. M. (1991). Feedback cancellation in hearing aids: Results from a computer simulation. *Journal of the Acoustical Society of America*, 39(3):553–562.

Kates, J. M. and Weiss, M. R. (1996). A comparison of hearing-aid array processing techniques. *Journal of the Acoustical Society of America*, 99(5):3138–3148.

Kates, J. M. (1999). Constrained adaptation for feedback cancellation in hearing aids. *Journal of the Acoustical Society of America*, 106(2):1010–1019.

Kates, J. M. (2000). Room reverberation effects in hearing aid feedback cancellation. *Journal of the Acoustical Society of America*, 109(1):367–378.

Kim, J. W., Seo, M. S., Yoon, B. S., Choi, S. I., and You, Y. G. (2000). A voice activity detection algorithm for wireless communication systems with dynamically varying background noise. *IEICE Transaction on Communication*, E83-B(2):414–418.

Kompis, M., Dillier, N., Francois, J., Tinembart, J., and Hausler, R. (1997). New target-signal-detection schemes for multi-microphone noise reduction systems for hearing aids. *IEEE Engineering in Medecine and Biology Society Conference (IEEE-EMBS)*, Chicago, USA:1990–1994.

Kompis, M. and Dillier, N. (2001a). 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.

Kompis, M. and Dillier, N. (2001b). 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.

Le Bouquin, R. and Faucon, G. (1995). Study of a voice activity detector and its influence on a noise reduction system. *Speech communication*, 16:245–254.

Lee, T. W. (1998). *Independent Component Analysis*. Kluwer Academic Publishers.

Lee, Y. C. and Ahn, S. S. (2001). An improved voice activity detection algorithm employing speech enhancement preprocessing. *IEICE Transaction on Fundamentals*, E84-A(6):1401–1405.

Leeuw, A. R. and Dreschler, W. A. (1991). Advantages of directional hearing aid microphones related to room acoustics. *Audiology*, 30(6):330–344.

Levitt, H., Neuman, A., Mills, R., and Schwander, T. (1986). A digital master hearing aid. *Journal of Rehabilitation Research and Development*, 23:79–87.

Li, M., McAllister, H. G., Black, N. D., and DePrez, T. A. (2001). Perceptual time-frequency subtraction algorithm for noise reduction in hearing aids. *IEEE Transaction on Biomedical Engineering*, BE-48(9):979–988.

Lim, J. S. and Oppenheim, A. V. (1979). Enhancement and bandwidth compression of noisy speech. *Proceedings of the IEEE*, 67(12):1586–1603.

Luo, F., Yang, J., Pavlovic, C., and Neuman, A. (2002). Adaptive null-forming scheme in digital hearing aids. *IEEE Transaction on Signals Processing*, SP-50(7):1583–1590.

Luts, H., Maj, J. B., Soede, W., and Wouters, J. (2003). Better speech perception in noise with an assistive multi-microphone array for hearing aids. *Accepted for publication in Ear and Hearing.*

Maj, J. B., Wouters, J., and Moonen, M. (2001a). A two-stage adaptive beamformer for noise reduction in hearing aids. *International Workshop on Acoustic Echo and Noise Control (IWAENC)*, September 10-13, Darmstadt, Germany:171–174.

Maj, J. B., Moonen, M., and Wouters, J. (2001b). Evaluation of an adaptive beamformer for behind-the-ear hearing aids. *Nederlands Akoestisch Genootschap (NAG)*, 157:61–70.

Maj, J. B., Wouters, J., and Moonen, M. (2002a). SVD-based optimal filtering technique for noise redcution in hearing aids using two microphones. *Journal on Applied Signal Processing*, 4:432–443.

Maj, J. B., Moonen, M., and Wouters, J. (2002b). Theoretical analysis of adaptive noise reduction algorithms for hearing aids. *European Signal Processing Conference (EUSIPCO)*, September 3-6, Toulouse, France:CD.

Maj, J. B., Moonen, M., and Wouters, J. (2002c). A robust voice activity detector for SVD-based noise reduction in hearing aids. *Signal Processing Symposium (SPS)*, March 21-22, Leuven, Belgium:189–192.

Maj, J. B., Wouters, J., and Moonen, M. (2003). Noise reduction results of an adaptive filtering technique for dual-microphone behind-the-ear hearing aids. *Accepted for publication in Ear and Hearing.*

Maj, J. B., Royackers, L., Moonen, M., and Wouters, J. (2003a). Comparison of adaptive noise reduction algorithms in dual microphone hearing aids. *International Workshop on Acoustic Echo and Noise Control (IWAENC)*, September 8-11, Kyoto, Japan:171–174.

Maj, J. B., Royackers, L., Wouters, J., and Moonen, M. (2003b). Comparison of adaptive noise reduction algorithms in dual microphone hearing aids. *Submitted to IEEE Transactions on Biomedical Engineering.*

Maj, J. B., Royackers, L., Moonen, M., and Wouters, J. (2003c). SVD-based optimal filtering for noise reduction in dual microphone hearing aids: A real time implementation and perceptual evaluation. *Submitted to IEEE Transactions on Biomedical Engineering.*

Maj, J. B., Royackers, L., Moonen, M., and Wouters, J. (2004). A real time implementation and an evaluation of an optimal filtering technique for noise reduction in dual microphone hearing aids. *Accepted for publication at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 17-21, Montreal, Canada.

Marzinzik, M. and Kollmeier, B. (2002). Speech pause detection for noise spectrum estimation by tracking power envelope dynamics. *IEEE Transactions on Speech and Audio Processing*, SAP-10(2):109–118.

Maxwell, J. A. and Zurek, P. M. (1995). Reducing acoustic feedback in hearing aids. *IEEE Transactions on Speech and Audio Processing*, SAP-3:304–313.

McAulay, R. J. and Malpass, M. L. (1980). Speech enhancement using a soft-decision noise suppression filter. *IEEE Transactions on Acoustics, Speech and Signals Processing*, ASSP-28:137–145.

Mittal, U. and Phamdo, N. (2000). Signal/noise KLT based approach for enhancing speech degraded by colored noise. *IEEE Transaction in Speech and Audio Processing*, SAP-8(2):159–167.

Miyoshi, M. and Kaneda, Y. (1988). Inverse filtering of room acoustics. *IEEE Transactions on Acoustics, Speech and Signals Processing*, ASSP-36(2):145–152.

Moonen, M., Van Dooren, P., and Vandewalle, J. (1991). A systolic algorithm for QSVD updating. *Signal Processing*, 25:203–213.

Moore, B. C. and Alkantara, J. I. (2001). The use of psychophysical tuning curves to explore dead regions in the cochlear. *Ear and Hearing*, 22:268–278.

NguyenThi, H. L. (1992). *Séparation aveugle de sources à bande large dans un mélange convolutif.* PhD thesis, Institut National Polytechnique de Grenoble.

Noordhoek, I. M. (2000). *Intelligibility of narrow-band speech and its relation to auditory functions in hearing impaired person.* PhD thesis, Vrije Universiteit Amsterdam.

Noordhoek, I. M., Housgast, T., and Festen, J. M. (2000a). Measuring the threshold for speech reception by adaptive variation of the signal bandwith. II. hearing impaired listeners. *Journal of the Acoustical Society of America*, 107(3):1685–1696.

Noordhoek, I. M., Housgast, T., and Festen, J. M. (2001). Relation between intelligibilty of narrow-band speech and auditory functions, both 1khz frequency region. *Journal of the Acoustical Society of America*, 109(3):1197–1212.

Ozer, H. and Tanyer, S. G. (1998). A geometric algorithm for a voice activity detection in nonstationary gaussian noise. *European Signal Processing Conference (EUSIPCO)*, September 8-11, Rhodes, Greece:CD.

Pavlovic, C. V. (1987). Derivation of premary parameters and procedures for use in speech intelligibility predictions. *Journal of the Acoustical Society of America*, 82(2):413–422.

Peterson, P. M., Durlach, N., Rabinowitz, W. M., and Zurek, P. M. (1987). Multimicrophone adaptive beamforming for interference reduction in hearing aids. *Journal of Rehabilitation Research and Development*, 24:103–110.

Peterson, P. M. (1989). *Adaptive array processing for multiple microphone hearing aids*. PhD thesis, MIT Cambridge.

Peutz, V. M. (1971). Articulation loss of consonants as a criterion for speech transmission in a room. *Journal of the Audio Engineering Society*, 19(11):915–919.

Plomp, R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. *Journal of the Acoustical Society of America*, 63(2):533–549.

Plomp, R. and Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology*, 18(1):43–52.

Plomp, R. (1986). A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *Journal of Speech and Hearing Research*, 29(2):146–154.

Plomp, R. (1988). The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function. *Journal of the Acoustical Society of America*, 83(6):2322–2327.

Rafaely, B., Roccasalva-Firenze, M., and Payne, E. (2000). Feedback path variability modeling for robust hearing aids. *Journal of the Acoustical Society of America*, 107(5):2665–2673.

Rezayee, A. and Gazor, S. (2001). An adaptive KLT approach for speech enhancement. *IEEE Transaction in Speech and Audio Processing*, SAP-9(2):87–95.

Ricketts, T. (2000a). Directivity quantification in hearing aids: fitting and measurement effects. *Ear and Hearing*, 21(1):45–58.

Ricketts, T. (2000b). The impact of head angle on monaural and binaural performance with directional and omnidirectional hearing aids. *Ear and Hearing*, 21(4):318–328.

Ricketts, T. and Henry, P. (2002). Evaluation of an adaptive, directional-microphone hearing aid. *International Journal of Audiology*, 41(2):100–112.

Rissanen, J. (1978). Modelling by shortest data description. *Automatica*, 14:465–471.

Rombouts, G. (2003). *Adaptive filtering algorithms for acoustic echo and noise reduction*. PhD thesis, Katholieke Universiteit Leuven.

Rosca, J., Balan, R., Ping, F. N., Beaugeant, C., and Gilg, V. (2002). Multichannel voice detection in adverse environments. *European Signal Processing Conference (EUSIPCO)*, September 3-6, Toulouse, France:CD.

Saunders, G. H. and Kates, J. M. (1997). Speech intelligibility enhancement using hearing-aid array processing. *Journal of the Acoustical Society of America*, 102(3):1827–1837.

Savoji, M. H. (1989). A robust algorithm for accurate endpointing of speech signals. *Speech Communication*, 8:45–60.

Schobben, D. (1999). *Efficient adaptive multi-channel concepts in acoustics: blind signal separation and echo cancellation*. PhD thesis, Technische Universiteit Eindhoven.

Schweitzer, C., Terry, M., and Grim, M. (1996). Three experimental measures of a digital beamforming signal processing algorithm. *Journal of American Academy of Audiology*, 7(4):230–239.

Schweitzer, C. (1997). Application of beamforming in a digital hearing system. *Seminars in Hearing*, 18(4):393–404.

Siqueira, M. G. and Alwan, A. (2000). Steady-state analysis of continuous adaptation in acoustic feedback reduction systems for hearing aids. *IEEE Transactions on Speech and Audio Processing*, SAP-8(4):443–453.

Soede, W., Berkhout, A. J., and Bilsen, F. A. (1993a). Development of a directional hearing instrument based on array technology. *Journal of the Acoustical Society of America*, 94(2):785–798.

Soede, W., Bilsen, F. A., and Berkhout, A. J. (1993b). Assessment of a directional microphone array for hearing-impaired listeners. *Journal of the Acoustical Society of America*, 94(2):799–808.

Sohn, J. and Sung, W. (1998). A voice activity detection employing soft decision based noise spectrum adaptation. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 12-15, Seattle, USA:365–368.

Sohn, J., Kim, N. S., and Sung, W. (1999). A statistical model based voice activity detection. *IEEE Signal Processing Letters*, 6(1):1–3.

Spriet, A., Moonen, M., and Wouters, J. (2001). Robustness analysis of GSVD based optimal filtering and generalized sidelobe canceller for hearing aid applications. *IEEE Workshop on Applications on Signal Processing to Audio and Acoustics (WASPAA 2001)*, October 21-24, New Paltz, USA:31–34.

Spriet, A., Moonen, M., and Proudler, I. (2002a). Feedback cancellation in hearing aids: an unbiased modeling approach. *European Signal Processing Conference (EUSIPCO 2002)*, September 3-6, Toulouse, France:CD.

Spriet, A., Moonen, M., and Proudler, I. (2002b). A multichannel subband GSVD approach to speech enhancement. *European Transactions on Telecommunications*, 13(2):149–158.

Spriet, A., Moonen, M., and Wouters, J. (2003a). The impact of speech detection errors on the noise reduction performance of multi-channel wiener filtering and generalized sidelobe cancellation. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 6-10, Hong Kong, China:501–504.

Spriet, A., Moonen, M., and Wouters, J. (2003b). Spatially preprocessed, speech distortion weighted multi-channel wiener filtering for noise reduction. *Submitted to Signal Processing.*

Spriet, A., Moonen, M., and Wouters, J. (2003c). Stochastic gradient based implementation of spatially preprocessed multi-channel wiener filter for noise reduction in hearing aids. *Submitted to IEEE Transactions on Signal Processing.*

Spriet, A., Moonen, M., and Wouters, J. (2004). Robustness analysis of multichannel wiener filtering and generalized sidelobe cancellation for multimicrophone noise reduction in hearing aids applications. *Accepted for publication in IEEE Transactions on Speech and Audio Processing.*

Stadler, R. W. and Rabinowitz, W. M. (1993). On the potential of fixed arrays for hearing aids. *Journal of the Acoustical Society of America*, 94(3):1332–1342.

Steiglitz, K. (1996). *A Digital Signal Processing Primer with Applications to Digital Audio and Computer Music.* Addition-Wesley.

Sydow, C. (1994). Broadband beamforming for a microphone array. *Journal of the Acoustical Society of America*, 96(2):845–849.

Tanyer, S. G. and Ozer, H. (2000). Voice activity detection in non-stationary noise. *Transaction on Speech and Audio Processing*, SAP-8(4):478–482.

Tchorz, J. (2001). Effects of microphone mismatch in directional hearing instruments. *Hearing Review*, 8(10):54–58.

Thomas, I. and Niederjohn, R. (1968). Enhancement of speech intelligibility at high noise levels by filtering and clipping. *Journal of the Audio Engineering Society*, 16:412–415.

Thompson, S. C. (1999). Dual microphones or directional-plus-omni: which is the best ? *The Hearing Review*, 3:31–35.

Thompson, S. C., Lo Presti, J. L., Ring, E. M., Nepomuceno, H. G., Beard, J. J., Ballad, W. J., and Carlson, E. V. (2002). Noise in miniature microphone. *Journal of the Acoustical Society of America*, 111(2):861–866.

Tucker, R. (1992). Voice activity detection using a periodicity measure. *IEE Proceedings*, 139(4):377–380.

Tyler, R. S. and Schum, D. J. (1995). *Assistive Devices for Persons with Hearing Impairment*. Allyn and Bacon.

Van Compernolle, D. (1990). Switching adaptive filters for enhancing noisy and reverberant speech from microphone array recordings. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 3-6, Albuquerque, USA:833–836.

Van Gerven, S. and Van Compernolle, D. (1995). Signal separation by symmetric adaptive adaptive decorrelation: stability, convergence and uniqueness. *IEEE Transaction on Signal Processing*, SP-43(7):1602–1612.

Van Gerven, S. (1996). *Adaptive noise cancellation and signal separation with applications to speech enhancement*. PhD thesis, Katholieke Universiteit Leuven.

Van Gerven, S. and Xie, F. (1997). A comparative study of speech detection methods. *Eurospeech*, September 22-25, Rhodes, Greece:1095–1098.

Vanden Berghe, J. and Happaerts, W. (1995). *Verbetering van spraakverstaanbaarheid bij hoorapparaten met behulp van adaptieve ruisonderdrukking in reele tijd*. Internal report, Katholieke Universiteit Leuven.

Vanden Berghe, J. and Wouters, J. (1998). An adaptive noise canceller for hearing aids using two nearby microphones. *Journal of the Acoustical Society of America*, 103(6):3621–3626.

Veit, I. and Sander, H. (1987). Production of spatially limited diffuse sound field in an anechoic room. *Journal of the Audio Engineering Society*, 35:138–142.

Versfeld, N., Daalder, L., Festen, J. M., and Houtgast, T. (2000). Extension of sentence materials for the measurement of the speech reception threshold. *Journal of the Acoustical Society of America*, 107(3):1671–1684.

Virag, N. (1999). Single channel speech enhancement based on masking properties of the human auditory system. *IEEE Transactions on Speech and Audio Processing*, SAP-7(2):126–137.

Weiss, M. (1974). Processing speech signals to attenuate interference. *IEEE Symposium on Speech Recognition*.

Welker, D. P., Greenberg, J. E., Desloge, J. G., and Zurek, P. M. (1997). Microphone-array hearing aids with binaural output- part 2: A two-microphone adaptive system. *IEEE Transactions on Speech and Audio Processing*, SAP-5(6):543–551.

Widrow, B. and Stearns, S. (1985). *Adaptive Signal Processing*. Englewood Cliffs.

Wouters, J., Damman, W., and Bosman, A. (1994). Vlaamse opname van woordenlijten voor spraakaudiometris. *Logopedie*, 7(6):28–33.

Wouters, J., Litiere, L., and van Wieringen, A. (1999). Speech intelligibility in noisy environments with one- and two- microphone hearing aids. *Audiology*, 38(2):91–98.

Wouters, J. and Vanden Berghe, J. (2001). Speech recognition in noise for cochlear implantees with a two- microphone monaural adaptive noise reduction system. *Ear and Hearing*, 22(5):420–430.

Wouters, J., Vanden Berghe, J., and Maj, J. B. (2002). Adaptive noise suppression for a dual microphone hearing aid. *International Journal of Audiology*, 41(7):401–407.

Yu, S. J. and Ueng, F. B. (2000). Blind adaptive beamforming based on generalized sidelobe canceller. *Signal Processing*, 80:2497–2506.

Appendix

A Signal processing basics

A brief review of discrete-time signal processing is presented and only the definitions which are important for the signal processing considered in this thesis are mentioned. For more information the readers are referred to the literature (Haykin, 1996; Steiglitz, 1996; Couch, 1997).

The recorded signals at the microphones of the digital hearing aids are equal to $u(k) = h(k) * s(k) + n(k)$, where the variables $u(k)$, $s(k)$ and $n(k)$ are discrete-time *stochastic processes*. A stochastic process is used to describe the time evolution of a statistical phenomenon according to probabilistic laws, however, it is not possible to predict exactly how the process evolves in time (e.g. speech signals, television signals, radar signals...).

The representation of a discrete-time stochastic process $x(k)$ is:

$$x(k) = \{\ldots \quad x_{-2} \quad x_{-1} \quad x_0 \quad x_1 \quad x_2 \quad \ldots\} \tag{A.1}$$

where the discrete time k is related to the actual time by $t = k/f_s$, through the sampling frequency f_s of the system. A stochastic process represents an infinite number of realizations of the process. One particular realization of a discrete-time stochastic process is called a *time series* and has a finite-duration sequence of length L. A time series of the discrete-time stochastic process $x(k) \in \mathbb{R}^L$ can be represented by a vector:

$$\mathbf{x}_k = [x_k \quad x_{k-1} \quad x_{k-2} \quad x_{k-L+1}]^T \tag{A.2}$$

where the superscript T denotes transposition. In practice, the characteristics of the discrete time variable $x(k)$ are estimated from the time series $\mathbf{x}_k$. A stochastic process is *stationary* if the statistical properties of several time series of the same stochastic process are invariant.

The vector $\mathbf{x}_k$ represents the discrete-time variable $x(k)$ in the *time domain* but can also be represented in the *frequency domain*. The representation of $\mathbf{x}_k$ in the frequency domain is obtained by applying the Discrete-Time Fourier Transform (DTFT) as:

$$\mathbf{x}(l) = \mathcal{F}\{\mathbf{x}_k\} = \sum_{k=0}^{L-1} \mathbf{x}_k e^{-j2\pi kl/L} \tag{A.3}$$

where the variable $l = 0, \quad 1, \quad 2, \quad \cdots, \quad L-1$ is related to the frequency by $f = l.f_s/L$ and $2\pi f$ is the angular frequency, lying in the Nyquist interval $-\pi < 2\pi f \leq \pi$.

The characteristics of a stochastic process are generally estimated by specifying the first (*mean function*) and the second (*autocorrelation function*) moments. The mean function of a *wide-sense stationary*[1] stochastic process $x(k)$ is defined as:

$$\mu(k) = \mathcal{E}\{x(k)\} \tag{A.4}$$

where $\mathcal{E}\{.\}$ denotes the statistical expectation operator. The autocorrelation function of the process is:

$$r_{x,x}(n) = \mathcal{E}\{x(k)x(k-n)^*\} \qquad n = 0, \quad \pm 1, \quad \pm 2, \quad \cdots \tag{A.5}$$

where the asterisk denotes complex conjugation. In the case of two jointly wide-sense stationary stochastic processes $x(k)$ and $y(k)$ ($y(k)$ is similarly defined as $x(k)$), the cross-correlation function is defined as:

$$r_{x,y}(n) = \mathcal{E}\{x(k)y(k-n)^*\} \qquad n = 0, \quad \pm 1, \quad \pm 2, \quad \cdots \tag{A.6}$$

In this thesis, the stochastic processes $x(k)$ and $y(k)$ are assumed *ergodic*[2]. This property allows one to compute the Power Spectral Density (PSD) $P_{x,x}(f)$ of $x(k)$ and the Cross-Power Spectral Density (C-PSD) $P_{x,y}(f)$ of $x(k)$ and $y(k)$. The $P_{x,x}(f)$ and the $P_{x,y}(f)$ are defined as the DTFTs of the correlation function $r_{x,x}(l)$ and the cross-correlation function $r_{x,y}(l)$, respectively:

$$\begin{aligned} P_{x,x}(f) &= \mathcal{F}\{r_{x,x}(n)\} \\ P_{x,y}(f) &= \mathcal{F}\{r_{x,y}(n)\} \end{aligned} \tag{A.7}$$

The *Complex Coherence Function* (CCF) between the signals $x(k)$ and $y(k)$ is defined as:

$$\Gamma_{x,y}(f) = \frac{P_{x,y}(f)}{\sqrt{P_{x,x}(f).P_{y,y}(f)}} \tag{A.8}$$

and the *Power Transfer Function* (PTF) $H_{x,y}$ between the time series $\mathbf{x}_k$ and $\mathbf{y}_k$, which can be considered for instance as an input and an output signal of a system, respectively, is defined as:

$$H_{x,y} = \frac{P_{y,y}(f)}{P_{x,x}(f)} \tag{A.9}$$

The correlation matrix $\mathbf{R}_{x,x} \in \mathbb{R}^{L \times L}$ of a stationary discrete-time stochastic process represented by a time series $\mathbf{x}_k$ is given as:

$$\mathbf{R}_{x,x} = \mathcal{E}\{\mathbf{x}_k \mathbf{x}_k^H\} \tag{A.10}$$

[1] A stationary process is stationary in the wide-sense if and only if the mean square of $x(k)$ has a finite value as $\mathcal{E}\{|x(k)|^2\} < \infty$.

[2] A stochastic process is ergodic if the mean-square value of the error between the ensemble average and the time average of a time series approaches zero.

where the superscript H denotes the Hermitian transposition. The correlation matrix can be expressed in the expanded form:

$$\mathbf{R}_{x,x} = \begin{pmatrix} r_{x,x}(0) & r_{x,x}(1) & \cdots r_{x,x}(L-1) \\ r^*_{x,x}(1) & r_{x,x}(0) & \cdots r_{x,x}(L-2) \\ \vdots & \vdots & \ddots \quad \vdots \\ r^*_{x,x}(L-1) & r^*_{x,x}(L-2) & \cdots \quad r_{x,x}(0) \end{pmatrix} \tag{A.11}$$

In a multiple microphone application, the vector $\mathbf{x}_k \in \mathbb{R}^{ML}$ takes the form:

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{x}_{1,k} & \mathbf{x}_{2,k} & \cdots & \mathbf{x}_{M,k} \end{bmatrix}^T \tag{A.12}$$

with

$$\mathbf{x}_{j,k} = \begin{bmatrix} x_{j,k} & x_{j,k-1} & \cdots & x_{j,k-L+1} \end{bmatrix}^T \tag{A.13}$$

where the j refers to the j-th microphone and M is the number of the microphones. The intermicrophone correlation matrix $\mathbf{R}_{x,x} \in \mathbb{R}^{ML \times ML}$ is then redefined as:

$$\mathbf{R}_{x,x} = \begin{pmatrix} \mathbf{R}_{1,1} & \mathbf{R}_{1,2} & \cdots & \mathbf{R}_{1,M} \\ \mathbf{R}^*_{1,2} & \mathbf{R}_{2,2} & \cdots & \mathbf{R}_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}^*_{1,M} & \mathbf{R}^*_{2,M} & \cdots & \mathbf{R}_{M,M} \end{pmatrix} \tag{A.14}$$

where $\mathbf{R}_{1,1}$ is the correlation matrix in the single microphone case (equation A.11).

B Linear algebra and matrix decomposition definitions

In this appendix some linear algebra and matrix decomposition definitions are briefly reviewed. A good reference text that focuses on numerical aspects and implementation issues is (Golub and Loan, 1996).

Definition A.1 Given the $N \times N$-dimensional real matrix $\mathbf{A}$, all (complex) N-dimensional vectors $\mathbf{v}_i \neq \mathbf{0}$ and (complex) scalars λ_i that satisfy

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i \tag{B.1}$$

are called *eigenvectors* and *eigenvalues* of the matrix $\mathbf{A}$.

Definition A.2 If the $N \times N$-dimensional real matrix $\mathbf{A}$ has N linearly independent eigenvectors, then the *eigenvalue decomposition* (EVD) of $\mathbf{A}$ is given by

$$\mathbf{A} = \mathbf{V}\mathbf{\Delta}\mathbf{V}^{-1} , \tag{B.2}$$

with $\mathbf{V}$ an $N \times N$-dimensional matrix containing the eigenvectors $\mathbf{v}_i$ as columns and $\mathbf{\Delta}$ a diagonal matrix containing the eigenvalues λ_i. Normally, the eigenvectors are normalized such that $||\mathbf{v}_i||_2 = 1$, $i = 1 \ldots N$. The EVD of a symmetric matrix $\mathbf{A}$ is equal to

$$\mathbf{A} = \mathbf{V}\mathbf{\Delta}\mathbf{V}^{T} , \tag{B.3}$$

with $\mathbf{V}$ an orthogonal matrix and all eigenvalues real scalars.

Definition A.3 Given the $P \times N$-dimensional real matrices $\mathbf{A}$ and $\mathbf{B}$, all (complex) N-dimensional vectors $\mathbf{x}_i \neq \mathbf{0}$ and (complex) scalars λ_i that satisfy

$$\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{B}\mathbf{x}_i \tag{B.4}$$

are called *generalized eigenvectors* and *generalized eigenvalues* of the matrices $\mathbf{A}$ and $\mathbf{B}$.

Definition A.4 If the $P \times N$-dimensional real matrices $\mathbf{A}$ and $\mathbf{B}$ have N linearly independent generalized eigenvectors, then the *generalized eigenvalue decomposition* (GEVD) of $\mathbf{A}$ and $\mathbf{B}$ is given by

$$\mathbf{A}\mathbf{X} = \mathbf{B}\mathbf{X}\mathbf{\Lambda} , \tag{B.5}$$

with $\mathbf{X}$ an $N \times N$-dimensional (invertible) matrix containing the generalized eigenvectors $\mathbf{x}_i$ as columns and $\mathbf{\Lambda}$ a diagonal matrix containing the generalized eigenvalues λ_i. Alternatively, the GEVD of $\mathbf{A}$ and $\mathbf{B}$ can be written as

$$\begin{cases} \mathbf{A} = \mathbf{Q}\mathbf{\Lambda}_A\mathbf{X}^{-1} \\ \mathbf{B} = \mathbf{Q}\mathbf{\Lambda}_B\mathbf{X}^{-1} , \end{cases} \tag{B.6}$$

with $\mathbf{Q}$ a $P \times N$ matrix and $\mathbf{\Lambda} = \mathbf{\Lambda}_A \mathbf{\Lambda}_B^{-1}$. The GEVD of two symmetric $N \times N$-dimensional matrices $\mathbf{A}$ and $\mathbf{B}$ is given by

$$\begin{cases} \mathbf{A} = \mathbf{Q}\mathbf{\Lambda}_A\mathbf{Q}^T \\ \mathbf{B} = \mathbf{Q}\mathbf{\Lambda}_B\mathbf{Q}^T \ , \end{cases} \tag{B.7}$$

with $\mathbf{Q}$ an invertible, but not necessarily orthogonal, matrix.

Definition A.5 The *singular value decomposition* (SVD) of the $P \times N$-dimensional real matrix $\mathbf{A}$ (with $P \geq N$) is given by

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \ , \tag{B.8}$$

with $\mathbf{U}$ a $P \times N$-dimensional orthogonal matrix containing the (left) singular vectors $\mathbf{u}_i$, $\mathbf{V}$ an $N \times N$-dimensional orthogonal matrix containing the (right) singular vectors $\mathbf{v}_i$, and $\mathbf{\Sigma}$ an $N \times N$-dimensional diagonal matrix containing the singular values σ_i, with $\sigma_1 \geq \sigma_2 \ldots \geq \sigma_N \geq 0$. The matrix $\mathbf{A}$ can be written as the dyadic decomposition

$$\mathbf{A} = \sum_{i=1}^{N} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \ , \tag{B.9}$$

such that

$$\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i, \quad \mathbf{A}^T\mathbf{u}_i = \sigma_i \mathbf{v}_i \ . \tag{B.10}$$

If the matrix $\mathbf{A}$ has rank $R < N$, then $N - R$ singular values are equal to zero, such that the SVD of $\mathbf{A}$ can be written as

$$\mathbf{A} = \mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}_1^T \ , \tag{B.11}$$

with $\mathbf{U}_1$ a $P \times R$-dimensional and $\mathbf{V}_1$ an $R \times R$-dimensional orthogonal matrix, and $\mathbf{\Sigma}_1 = \text{diag}\{\sigma_1, \ldots, \sigma_R\}$ an $R \times R$-dimensional diagonal matrix.

Definition A.6 The (Moore-Penrose) *pseudo-inverse* of the $P \times N$-dimensional real matrix $\mathbf{A}$ with rank R is defined using (B.11) as

$$\mathbf{A}^\dagger = \mathbf{V}_1 \mathbf{\Sigma}_1^{-1} \mathbf{U}_1^T \ . \tag{B.12}$$

Definition A.7 The *generalized singular value decomposition* (GSVD) of the $P \times N$-dimensional real matrix $\mathbf{A}$ and the $Q \times N$-dimensional real matrix $\mathbf{B}$ (with $P \geq N$ and $Q \geq N$) is given by

$$\begin{cases} \mathbf{A} = \mathbf{U}_A \mathbf{\Sigma}_A \mathbf{Q}^T \\ \mathbf{B} = \mathbf{U}_B \mathbf{\Sigma}_B \mathbf{Q}^T \ , \end{cases} \tag{B.13}$$

with $\mathbf{U}_A$ a $P \times N$-dimensional orthogonal matrix, $\mathbf{U}_B$ a $Q \times N$-dimensional orthogonal matrix, $\mathbf{Q}$ an $N \times N$-dimensional invertible, but not necessarily

orthogonal, matrix containing the (right) generalized singular vectors $\mathbf{q}_i$ and $\mathbf{\Sigma}_A$ and $\mathbf{\Sigma}_B$ $N \times N$-dimensional diagonal matrices containing σ_{Ai} and σ_{Bi}, with $\sigma_{A1} \geq \sigma_{A2} \ldots \geq \sigma_{AN} \geq 0$ and $0 \leq \sigma_{B1} \leq \sigma_{B2} \ldots \leq \sigma_{BN}$. The generalized singular values are equal to σ_{Ai}/σ_{Bi}. The matrices $\mathbf{A}$ and $\mathbf{B}$ can be written as the dyadic decomposition

$$\begin{cases} \mathbf{A} = \sum_{i=1}^{N} \sigma_{Ai} \mathbf{u}_{Ai} \mathbf{q}_i^T \\ \mathbf{B} = \sum_{i=1}^{N} \sigma_{Bi} \mathbf{u}_{Bi} \mathbf{q}_i^T \ . \end{cases} \tag{B.14}$$

If the matrix is defined as $\mathbf{T} = \mathbf{Q}^{-T}$, containing the vectors $\mathbf{t}_i$, such that $\mathbf{q_i}^T \mathbf{t}_j = \delta_{ij}$, with δ_{ij} the Kronecker-delta, then

$$\begin{cases} \mathbf{A}\mathbf{t}_i = \sigma_{Ai} \mathbf{u}_{Ai}, & \mathbf{A}^T \mathbf{u}_{Ai} = \sigma_{Ai} \mathbf{q}_i \\ \mathbf{B}\mathbf{t}_i = \sigma_{Bi} \mathbf{u}_{Bi}, & \mathbf{B}^T \mathbf{u}_{Bi} = \sigma_{Bi} \mathbf{q}_i \ . \end{cases} \tag{B.15}$$

Definition A.8 The QR-decomposition of the $P \times N$ real matrix $\mathbf{A}$ (with $P \geq N$) is defined

$$\mathbf{A} = \mathbf{Q}_A \mathbf{R}_A \ , \tag{B.16}$$

with $\mathbf{Q}_A$ a $P \times N$ orthogonal matrix and $\mathbf{R}_A$ an $N \times N$ upper-triangular matrix.

List of Publications

Publications in International Journals

Jean-Baptiste Maj, Liesbeth Royackers, Marc Moonen, Jan Wouters: IEEE Transactions on Biomedical Engineering. 'SVD-based optimal filtering for noise reduction in dual microphone hearing aids: a real time implementation and perceptual evaluation'. Submitted.

Jean-Baptiste Maj, Liesbeth Royackers, Jan Wouters, Marc Moonen: IEEE Transactions on Biomedical Engineering. 'Comparison of adaptive noise reduction algorithms in dual microphone hearing aids'. Submitted.

Jean-Baptiste Maj, Marc Moonen, Jan Wouters: Ear&Hearing. 'Noise reduction results of an adaptive filtering technique for behind-the-ear hearing aids with two microphones'. Accepted for publication.

Heleen Luts, Jean-Baptiste Maj, Wim Soede, Jan Wouters: Ear&Hearing. 'Evaluation of a fixed beam array for speech perception in noise'. Accepted for publication.

Jan Wouters, Jeff Vanden Berghe, Jean-Baptiste Maj: International Journal of Audiology. 'Adaptive noise suppression for a dual microphone hearing aid'. Volume 41, No.7, pp.401-407, October 2002.

Jean-Baptiste Maj, Marc Moonen, Jan Wouters: Journal on Applied Signal Processing. 'SVD-based optimal filtering technique for noise reduction in hearing aids using two microphones'. Volume 2002, n^{o}4, April 2002, pp.432-443.

Publications in National Journals

Jan Wouters, Luc Geurts, Johan Laneau, Heleen Luts, Jean-Baptiste Maj, Marc Moonen, Liesbeth Royackers, Ann Spriet, Tim Van den Bogaert, Astrid van Wieringen: Acta oto-rhino-laryngologica belgica. 'Digital hearing aids and future directions for hearing aids'. Volume 56, pp. 357-361, 2002.

Publications in Proceedings of International Conferences

Jean-Baptiste Maj, Marc Moonen, Jan Wouters: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). 'A real time implementation and an evaluation of an optimal filtering technique for noise reduction in dual microphone hearing aids'. Montréal, Canada, May 17th-21st, 2004, Accepted for publication.

Jean-Baptiste Maj, Marc Moonen, Jan Wouters: IEEE International Workshop on Acoustic Echo and Noise Control (IWAENC). 'Comparison of adaptive noise reduction algorithms in dual microphone hearing aid'. Kyoto, Japan, September 8th-11th, 2003, pp. 171-174.

Jean-Baptiste Maj, Marc Moonen, Jan Wouters: European Signal Processing Conference (EUSIPCO). 'Theoretical analysis of adaptive noise reduction algorithms for hearing aids'. Toulouse, France, September 3rd-6th, 2002, CD.

Jean-Baptiste Maj, Marc Moonen, Jan Wouters: Signal Processing Symposium (SPS). 'A robust voice activity detector for a SVD-based technique for noise reduction in hearing aids'. Leuven, Belgium, March 21st-22sd, 2002, pp.189-192.

Jean-Baptiste Maj, Jan Wouters, Marc Moonen: NAG-NVA studiedag. 'Evaluation of an adaptive beamformer for behind-the-ear hearing aids'. Utrecht, The Nederlands, September 28th, 2001, pp.61-70.

Jean-Baptiste Maj, Jan Wouters, Marc Moonen: IEEE International Workshop on Acoustic Echo and Noise Control (IWAENC). 'A two-stage adaptive beamformer for noise reduction in hearing aids'. Darmstadt, Germany, September 10th-13rd, 2001, pp. 171-174.

Jean-Baptiste Maj, Marc Moonen, Jan Wouters: Signal Processing Symposium (SPS). 'A comparison of different methods for noise reduction in hearing aids'. Hilvarenbeek, The Nederlands, March 23rd-24th, 2000, CD.

Olivier Caspary, Jean-Baptiste Maj: Groupe de Recherche En Traitement du Signal et Image (GRETSI). 'Détection de sinusoides noyées dans du bruit par les régions d'inclusion des valeurs propres'. Vannes, France, September 13th-17th, 1999, CD.

Contributions to Conferences

Jean-Baptiste Maj, Liesbeth Royackers, Jan Wouters, Marc Moonen: Symposium of the Belgian Society for Audiology. 'A first perceptual evaluation of an optimal filtering technique for noise reduction in dual microphone hearing aids'. Liège, Belgium, December 13th, 2003.

Jean-Baptiste Maj, Marc Moonen, Jan Wouters: Belgian Society for Audiology. 'Noise reduction results of an adaptive beamformer for behind-the-ear hearing aids'. Spa, Belgium, September 29th, 2001.

Jean-Baptiste Maj, Marc Moonen, Jan Wouters: XXV International Congress of Audiology. 'Field trial of a two-stage adaptive noise reduction strategy for dual microphone hearing aids'. Den Haag, The Nederlands, August 27th-31st, 2000.

Jean-Baptiste Maj, Marc Moonen, Jan Wouters: Belgian Society for Audiology. 'Improvement of the speech intelligibility in a noisy environment : a summary of different methods'. Affligem, Belgium, June 6th, 1999.

Curriculum Vitae

Jean-Baptiste Maj was born in Dijon, France, on October 8th 1973. In 1993, he graduated from the Ecole Spéciale de Radioelectricité in Nancy (France). Between 1993 and 1998 he studied Applied Sciences at the University of Nancy (Université Henri Poincaré), where he received his DEA degree in Electrical Engineering. In April 1999, he started research (Ph.D. student) at the Electrical Engineering Department and the Laboratory for Experimental ORL of the Katholieke Universiteit Leuven, Leuven, Belgium.