KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Bernardi G (2018), **Design and Evaluation of Feedback Control Algorithms for Implantable Hearing Devices** PhD thesis, Faculty of Engineering, KU Leuven (Leuven, Belgium), Feb. 2018, 206 p.
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	Klik hier als u tekst wilt invoeren.
Journal homepage	Klik hier als u tekst wilt invoeren.
Author contact	giuliano.bernardi@esat.kuleuven.be *+ 32 (0)16 321797*
IR	https://lirias.kuleuven.be/handle/123456789/604989

(article begins on next page)

KU LEUVEN

ARENBERG DOCTORAL SCHOOL
FACULTY OF ENGINEERING SCIENCE
DEPARTMENT OF ELECTRICAL ENGINEERING

Design and Evaluation of Feedback Control Algorithms for Implantable Hearing Devices

Giuliano Bernardi

Supervisor:
Prof. dr. ir. Marc Moonen
Co-supervisors:
Prof. dr. Jan Wouters
Prof. dr. ir. Toon van Waterschoot

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

February 2018

Design and Evaluation of Feedback Control Algorithms for Implantable Hearing Devices

Giuliano BERNARDI

Examination committee:
Prof. dr. Adhemar Bultheel, chair
Prof. dr. ir. Marc Moonen, supervisor
Prof. dr. Jan Wouters, co-supervisor
Prof. dr. ir. Toon van Waterschoot, co-supervisor
Prof. dr. ir. Sabine Van Huffel
Prof. dr. ir. Hugo Van hamme
Prof. dr. Ian Proudler
(Loughborough University)
Prof. dr. Nicolas Verhaert
(KU Leuven, UZ Leuven)

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

February 2018

© 2018 KU Leuven – Faculty of Engineering Science
Uitgegeven in eigen beheer, Giuliano Bernardi, Kasteelpark Arenberg 10, B-3001 Leuven (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, 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, electronic or any other means without written permission from the publisher.

Preface

I still recall what I said when I first started to ponder about pursuing a Ph.D. It was right after finishing my master's thesis in Denmark, while talking with a friend: "That doesn't sound so bad. It's like working on a new 3-year-long master's thesis... and they even pay you!". However, as in all estimation problems, things might get a bit out of hand if one does not account for the variance of the estimate. So here am I, more than five years since I started, at the end of this journey. A journey that has, admittedly, been a very rewarding one, from the educational, the working, and the social point of view. Therefore, the least I can do is to thank all the people who walked with me, supported and helped me through this *unforgettablePhDjourney*™.

I would like to start by thanking my supervisors, i. e. those people that more directly helped me reaching the end of this Ph.D.

First, I would like to thank my main supervisor, Prof. Marc Moonen, for the opportunity he gave me[1] and for the great supervision he has provided throughout the years. Besides his academic qualities that the reader can easily find online, I really enjoyed his sense of humor that made our meetings (and the several "road trips" to Mechelen) an occasion to both learn and have fun.

Next, I would like to thank my co-supervisor Prof. Jan Wouters, for the fruitful discussions about those parts of my work more closely related to the audiology world. Additionally, I would like to thank Jan for introducing me to the very nice hearing-aids conferences (both content-wise and landscape-wise) held in Lake Tahoe.

Finally, I would like to thank my (daily) co-supervisor Prof. Toon van Waterschoot. His expertise, enthusiasm, help, suggestions, and ideas have all been necessary elements for the completion of this Ph.D. Additionally, his

[1] Despite the fact that I showed up at the interview with a couple of friends (I know it was very unprofessional of me but SNCB/NMBS heavily disrupted my plan) and that I do not play any musical instrument.

funny stories and cheerful attitude made him a great colleague throughout the years.

Besides my supervisors, I would like to thank the chair and members of the examination committee for their time and their helpful feedback on my thesis. Additionally, I would like to separately thank the members of my supervisory committee, Prof. Sabine Van Huffel and Prof. Hugo Van hamme, for their time and their helpful feedback throughout the different milestones of my Ph.D.

Next come the colleagues of the DSP group. Thanks Rodrigo, Joe, Bruno, Pepe, Paschalis, Geert, Alex, and Johnny, the *old school* colleagues who helped me with the transition into the academic world and passed on the legendary stories that took place in the group before my arrival. Thanks to the colleagues who more recently left the DSP group: Enzo, Marijn, Wouter B., Martijn, David, Jorge, and Hassan. And thanks to all my current great colleagues: Amin, Rodolfo, Giacomo, Niccoló, Hanne, Fernando, Mina, Thomas, Jeroen, Wouter L., Neetha, Filippos, and Robbe. I could say that throughout my Ph.D. I have been happy to come to ESAT just because of my work or because of the beautiful scenery, but that would only be a partial truth. The social side of coming to ESAT, e. g. the discussions at lunch, the coffee breaks, the Wikipedia knowledge spreads, and, naturally, the nice work related discussions made the Ph.D. experience a whole lot better!

I am also grateful to the external people that followed my project, both from Cochlear and ExpORL, and helped me multiple times. Thanks Bas, Alberto, Joris, Federico, Martin, Tom, and Michael.

And in the final part of this academic thanksgiving list, thanks to Till Tantau and Christian Feuersänger for inventing PGF/Ti*k*Z, both a blessing and a curse for everyone who loves high quality graphics and typesetting. I am completely aware of the fact that I will probably have to ditch PGF/Ti*k*Z as soon as I leave academia. For the time being, however, (academic) life is too short to have lousy graphics.

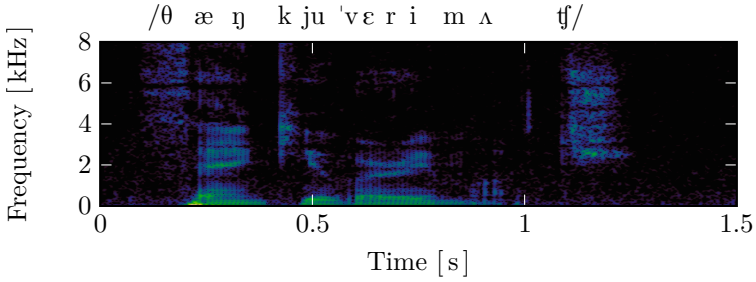
Thanks to the Brusselsestraat-19 team and friends Daniele, Oreste, and Leo for all the good times we have spent together. Thanks to the people from the Salcazzo crew and all the Leuven friends I have not mentioned yet: Gabriele, Martina, Carlo, Francesco, Enrico, Federico, Alessandra, Silvia, Ece (Işıl for the friends), Alice, Ana, Marta, Dan, Nina, Baharak, and Maria.

Even though I was planning to have completely non-overlapping partitions in this section, I cannot avoid to mention once more those friends I like to think to as my *Belgian family*: thanks Oreste, Giacomo, Ece, Leo, and Deniz for the countless moments spent together having fun, eating and drinking, watching South Park, and animatedly discussing about some random fact read online.

Un gran ringraziamento va anche alla mia famiglia: ai miei carissimi genitori, Duilio e Luigina, ed ai miei fratelli per avermi sempre sostenuto (sia moralmente che finanziariamente) in un percorso formativo molto poco convenzionale in quel di Pagnano d'Asolo e nonostante le difficoltà derivanti dal vivere a 1200 km di distanza. Grazie a Gigi, Jessica, Massimo, Erica, Stefano, Simone, Cristina, Matteo, Isaia, Andrea, Federico, e Chiara, ossia gli amici pre-Belgio da Pagnano e dintorni, per farmi sentire a casa ogni volta che ritorno. Grazie a Marcomardegan™, Angela, Giorgia, Massimo B., Michele, Giulio B., Giulio F., Giulia, Michela, Marco, Gabriele e gli altri *veggghi* dell'UniPD.

Thanks Bo, Alex, Fede, Edd, Claudia, Rok, Martin, Fernando, and Andreas, i. e. the great friends I made while I was in Denmark. I do not get to see you as much as I would like to anymore due to the divergence of our career paths, but I always enjoy spending time with you.

Finally, a special thanks goes to Deniz. Thank you, *tatlım*, for the great time spent together, for all the laughs in the funny and silly moments, for the *lezzetli* Turkish dishes you made me discover, for your love and support [especially during the tough times of my Ph.D. (years 3 to 4, I am looking at you!)], for teaching me about epigenetics and carbon nanotubes (☺), and for much more. Here is a nice heart ideograph ❤ or, given your research expertise, a more scientifically-sound heart drawing  for you.[2] On a side note, the language enthusiast in me also wants to thank you for introducing me to a very beautiful non-Indoeuropean language; agglutination is a lot of fun, from an engineering perspective! *Bu harika fırsat için, çok teşekkür ederim!*



Giuliano Bernardi
February 2018

[2] Unfortunately, it is not my work. The full code and an interesting (!?) discussion on how to draw hearts in Ti*k*Z can be found at `https://tex.stackexchange.com/questions/139733/`

Abstract

Using a hearing device is one of the most successful approaches to partially restore the degraded functionality of an impaired auditory system. However, due to the complex structure of the human auditory system, hearing impairment can manifest itself in different ways and, therefore, its compensation can be achieved through different classes of hearing devices.

Although the majority of hearing devices consists of conventional hearing aids (HAs), several other classes of hearing devices have been developed. For instance, bone-conduction devices (BCDs) and cochlear implants (CIs) have successfully been used for more than thirty years. More recently, other classes of implantable devices have been developed such as middle ear implants (MEIs), implantable BCDs, and direct acoustic cochlear implants (DACIs). Most of these different classes of hearing devices rely on a sound processor running different algorithms able to compensate for the hearing impairment. Nowadays, fully digital sound processors are the norm and this allows the use of advanced algorithms to tackle the different issues a hearing device might have to compensate for. Examples of algorithms implemented in a sound processor are, among others, noise reduction, nonlinear compression, sound scene classification, binaural enhancement, and, most importantly for the scope of this thesis, feedback cancellation.

In a hearing device, feedback arises when a coupling exists between the output (i. e. the loudspeaker or a different kind of actuator) and the input (i. e. the microphone). Feedback in hearing devices gives rise to different kinds of acoustic artifacts and sound degradation which can perceptually be very annoying. Therefore, several approaches to tackle this problem have been proposed. With the advent of digital hearing devices, approaches attempting to reduce feedback through adaptive algorithms have become more viable and are, nowadays, widespread. However, despite the advances of the recent years, the feedback problem in hearing devices has not yet been solved. This is due to different reasons such as, among others, the existence of different classes and kinds of hearing devices requiring specific algorithmic tailoring, the presence

of strong power constraints requiring low-complexity algorithms, and the great flexibility requirements a feedback control algorithm must have in order to cope with different daily life activities and soundscapes.

This thesis presents three different tasks related to the development of a feedback control strategy for a novel hearing device as follows: 1) the feedback characterization in two novel hearing devices; 2) the presentation of new algorithms for feedback control and their comparison with the state of the art; 3) the subjective and objective evaluation of the developed algorithms in terms of sound quality.

The introductory chapter provides a brief description of the auditory system and of the concept of hearing loss. Subsequently, the main differences between six classes of hearing devices are explained. Finally, an account of the feedback problem in hearing devices is given by discriminating between feedback in conventional HAs and feedback in implantable hearing devices.

The first part of this thesis describes the data collection and the analysis of a series of feedback characterization measurements for two novel implantable hearing devices, the Cochlear™ Codacs™ DACI and an early prototype of a bone conduction implant concept (BCIC). The measurements have been performed on fresh frozen cadaver heads and are used to investigate different important aspects of the feedback these two implants may experience, such as specimen-dependent behaviors, nonlinearities, and effects of structure-borne mechanical versus acoustic feedback.

The second part of this thesis introduces and describes two novel adaptive feedback cancellation (AFC) algorithms providing either comparable or better performance than existing algorithms, and both based on the prediction-error method (PEM) method. The first is an all-frequency-domain method, i. e. the frequency-domain prediction-error-method-based adaptive filter (FD-PEMAF), relying solely on frequency-domain signal-processing operations. The second is a PEM-based AFC algorithm replacing the standard adaptive filter with a simplified Kalman filter, i. e. the PEM-based frequency-domain Kalman filter (PEM-FDKF).

The third part of this thesis describes a study based on a subjective listening test to assess the sound-quality degradation caused by different AFC algorithms, showing a lower sound-quality degradation compared to existing algorithms, introduced by the PEM-FDKF. Additionally, the subjective listening test results are compared to the sound quality predicted by a batch of different objective measures.

Finally, the contributions of this thesis are reiterated and possible future research directions are explored.

Korte Inhoud

Eén van de meest succesvolle oplossingen om het verminderd functioneren van het menselijk gehoorsysteem gedeeltelijk te herstellen bestaat in het gebruik van een hoortoestel. De complexe structuur van het gehoorsysteem impliceert echter dat gehoorverlies zich op verscheidene manieren kan manifesteren, en om die reden kan het herstel ervan met verschillende soorten hoortoestellen worden bereikt.

Hoewel conventionele hoorapparaten (HAs) nog steeds de meest gebruike hoortoestellen zijn, werden er ook verscheidene andere soorten hoortoestellen ontwikkeld. Zo worden bijvoorbeeld hoortoestellen met beengeleiding (BCDs) en cochleaire implantaten (CIs) reeds meer dan dertig jaar met succes gebruikt. Recent werden nog andere soorten implanteerbare hoortoestellen ontwikkeld zoals middenoorimplantaten (MEIs), implanteerbare BCDs en direct akoestische cochleaire implantaten (DACIs). De meeste van deze verschillende soorten hoortoestellen zijn gebaseerd op een geluidsprocessor waarin verschillende algoritmes draaien om het gehoorverlies te compenseren. Vandaag de dag zijn volledig digitale geluidsprocessoren de norm, waardoor geavanceerde algoritmes kunnen worden gebruikt die de verschillende problemen kunnen aanpakken die in het hoortoestel moeten worden opgelost. Voorbeelden van algoritmes die in de geluidsprocessor geïmplementeerd worden, zijn o.a. ruisonderdrukking, niet-lineaire compressie, classificatie van geluidsscènes, binaurale signaalverbetering en feedbackonderdrukking, wat het belangrijkste thema van dit proefschrift vormt.

In een hoortoestel ontstaat feedback wanneer er een terugkoppeling bestaat tussen de uitgang (i.e. de luidspreker of een ander soort actuator) en de ingang (i.e. de microfoon). Feedback in hoortoestellen veroorzaakt verschillende soorten akoestische artefacten en geluidsdegradatie wat tot een een erg onaangename geluidswaarneming kan leiden. Om die reden zijn verschillende methodes ontwikkeld om dit probleem aan te pakken. Met de komst van digitale hoortoestellen werden methodes voor het onderdrukken van feedback

d.m.v. adaptieve algoritmes performanter en daardoor worden ze alsmaar vaker toegepast. Toch is het feedbackprobleem in hoortoestellen ondanks de vooruitgang in de laatste jaren nog niet opgelost. Daar zijn verschillende oorzaken voor zoals o.a. het feit dat er verschillende soorten en types hoortoestellen bestaan die specifieke algoritmische aanpassingen vereisen, de eis voor een beperkt vermogenverbruik en dus een lage algoritmische complexiteit, en de hoge graad van flexibiliteit die een feedbackbeheersingsalgoritme moet vertonen om te functioneren in uiteenlopende omstandigheden qua gebruik en akoestiek.

In dit proefschrift worden drie taken voorgesteld die deel uitmaken van de ontwikkeling van een feedbackbeheersingsstrategie voor een nieuw hoortoestel: 1) de karakterisering van feedback in twee nieuwe hoortoestellen; 2) de uiteenzetting van nieuwe algoritmes voor feedbackbeheersing en de vergelijking ervan met de state of the art; 3) de subjectieve en objectieve evaluatie van de ontwikkelde algoritmes in termen van geluidskwaliteit.

Het inleidende hoofdstuk bevat een korte beschrijving van het gehoorsysteem en het concept van gehoorverlies. Vervolgens worden de belangrijkste verschillen tussen zes soorten hoortoestellen uitgelegd. Tot slot wordt een overzicht gegeven van het feedbackprobleem in hoortoestellen door een onderscheid te maken tussen feedback in conventionele HAs en in implanteerbare hoortoestellen.

Het eerste deel van dit proefschrift beschrijft de datacollectie en de analyse van een reeks metingen om feedback te karakteriseren voor twee nieuwe implanteerbare hoortoestellen, met name de Cochlear Codacs DACI en een vroeg prototype van een implanteerbaar concept met beengeleiding (BCIC). De metingen werden uitgevoerd op vers ingevroren kadaverhoofden en worden gebruikt om verschillende belangrijke aspecten van feedback te onderzoeken die in deze twee implantaten kunnen voorkomen, zoals specimenafhankelijk gedrag, niet-lineariteiten, en effecten van structuurgedragen mechanische versus akoestische feedback.

Het tweede deel van dit proefschrift introduceert en beschrijft twee nieuwe algoritmes voor adaptieve feedbackonderdrukking (AFC) die vergelijkbaar of beter presteren dan bestaande algoritmes en beide gebaseerd zijn op de predictiefoutmethode (PEM). Het eerste algoritme is een volledig frequentiedomeingebaseerde methode, i.e. het frequentiedomein-predictiefoutgebaseerd adaptief filter (FD-PEMAF), die enkel signaalverwerkingsoperaties in het frequentiedomein uitvoert. Het tweede algoritme is een PEM-gebaseerd AFC-algoritme waarin het standaard adaptief filter wordt vervangen door een vereenvoudigd Kalmanfilter, i.e. het PEM-gebaseerd frequentiedomein-Kalmanfilter (PEM-FDKF).

Het derde deel van dit proefschrift beschrijft een studie gebaseerd op een subjectieve luistertest om de degradatie van geluidskwaliteit, veroorzaakt door verschillende AFC-algoritmes, te evalueren. Uit deze studie blijkt dat het PEM-FDKF een beperkte degradatie van geluidskwaliteit veroorzaakt in vergelijking met bestaande algoritmes. Daarnaast worden de resultaten van de subjectieve luistertest vergeleken met de geluidskwaliteit zoals voorspeld door een reeks objectieve maten.

Tot slot worden de bijdragen van dit proefschrift samengevat en worden mogelijke richtingen voor toekomstig onderzoek uitgestippeld.

Glossary

Abbreviations

cf.	*conferre*, compare with, see also
e. g.	*exempli gratia*, for example
i. e.	*id est*, that is
i.i.d.	independent and identically distributed
w.r.t.	with respect to

Acronyms

ε-RMSE	epsilon-insensitive root-mean-square error
AB	airborne
AB1	airborne1
AB2	airborne2
AEC	acoustic echo cancellation
AFC	adaptive feedback cancellation
AIR	acoustic impulse response

AMSL acoustical-to-mechanical sensitivity level

ANOVA analysis of variance

AP affine projection

AR autoregressive

ARMA autoregressive-moving-average

ASG added stable gain

ATF actuator transfer function

BAHA bone-anchored hearing aid

BCD bone-conduction device

BCI bone-conduction implant

BCIC bone conduction implant concept

BNLMS block normalized least mean squares

BTE behind-the-ear

CHL conductive hearing loss

CI cochlear implant

DACI direct acoustic cochlear implant

DFT discrete Fourier transform

DIR deconvolved impulse response

DSP digital signal processor

DTR deconvolved total response

ESS exponential sine sweep

FD-PEMAF frequency-domain prediction-error-method-based adaptive filter

FDAF frequency-domain adaptive filter

FDKF frequency-domain Kalman filter

HA hearing aid

HAAQI HA audio quality index

HASPI HA speech intelligibility index

HASQI HA speech quality index
HI hearing impaired

IC identifiability condition
IR impulse response
ITU International Telecommunications Union

LMS least mean squares
LS least squares

MEI middle ear implant
MES middle ear surgery
MHL mixed hearing loss
Mis misadjustment
MMSE minimum mean-square error
MS measurement session
MSE mean-square error
MSG maximum stable gain
MUSHRA Multi Stimulus Tests with Hidden Reference and Anchor

NH normal hearing
NLMS normalized least mean squares

OLA overlap and add
OLS overlap and save
OTF output transfer function

PA public address
PB partitioned block
PBCD percutaneous bone-conduction device
PBFD partitioned-block frequency-domain
PBFDAF partitioned-block frequency-domain adaptive filter
PBFDKF partitioned-block frequency-domain Kalman filter

PE prediction error

PEAQ perceptual evaluation of audio quality

PEM prediction-error method

PEM-AFC PEM-based adaptive feedback cancellation

PEM-AFROW PEM-based adaptive filtering with row operations

PEM-FDAF PEM-based frequency-domain adaptive filter

PEM-FDKF PEM-based frequency-domain Kalman filter

PEM-PBFDAF PEM-based partitioned-block frequency-domain adaptive filter

PEM-PBFDKF PEM-based partitioned-block frequency-domain Kalman filter

PEMO-Q perceptual-model-based quality prediction method

PESQ perceptual evaluation of speech quality

PNLMS proportionate NLMS

PSD power spectral density

PSM perceptual similarity measure

RF radio-frequency

RIR room impulse response

RLS recursive least squares

RMSE root-mean-square error

SD frequency-weighted log-spectral signal distortion

SNHL sensorineural hearing loss

SNR signal-to-noise ratio

STFT short-time Fourier transform

STOI short-time objective intelligibility

TBCD transcutaneous bone-conduction device

TF transfer function

THD total harmonic distortion

uFD-PEMAF unconstrained FD-PEMAF

Mathematical notation

Fixed notation

a	scalar
$\mathbf{a}$	vector
$\mathbf{A}$	matrix
$.^*$	scalar conjugate transpose
$.^{\mathrm{T}}$	vector or matrix transpose
$.^{\mathrm{H}}$	vector or matrix conjugate transpose
$.^{-1}$	matrix inverse
$\mathrm{diag}\{\cdot\}$	maps an $M \times 1$ vector to the diagonal of an $M \times M$ diagonal matrix or maps an $M \times M$ matrix to the $M \times 1$ vector given by its diagonal
$\mathbb{E}\{\cdot\}$	mathematical expectation
$\min_x$	minimize over x
$\arg\min_x$	argument of the minimum
$\mathbb{C}$	set of complex numbers
$\mathbb{N}$	set of natural numbers
$\mathbb{R}$	set of real numbers
$\mathbb{Z}$	set of integer numbers
$\forall$	for all
$\in$	set membership
$\lvert\cdot\rvert$	cardinality of a set or absolute value
$\lvert\lvert\cdot\rvert\rvert$	norm
$\lVert\mathbf{a}\rVert_{\mathbf{A}}$	norm of $\mathbf{a}$ weighted by the positive definite matrix $\mathbf{A}$
$\approx$	approximately equal to
$\gg$	much greater than
$\ll$	much less than
f_s	sampling frequency
q^{-1}	time delay operator, i. e. $q^{-k}u(t) = u(t-k)$

Chapter 1 notation

t	discrete time index
ω	angular frequency
$d[t, \hat{\mathbf{f}}(t)]$	time-domain error signal
$u(t)$	time-domain loudspeaker signal
$v(t)$	time-domain source signal
$x(t)$	time-domain feedback signal
$y(t)$	time-domain microphone signal
$\hat{y}[t\|\hat{\mathbf{f}}(t)]$	time-domain predicted feedback signal
$F_{\mathrm{t}}(q,t)$	true feedback path
$\hat{F}(q,t)$	estimated feedback path model
$G(q,t)$	electroacoustic forward path
$A(\omega)$	actuator frequency response
$F(\omega)$	total feedback path frequency response
$F_{\mathrm{A}}(\omega)$	acoustic feedback path frequency response
$F_{\mathrm{E}}(\omega)$	electrical feedback path frequency response
$F_{\mathrm{M}}(\omega)$	magnetic feedback path frequency response
$F_{\mathrm{SBM}}(\omega)$	structure-borne mechanical feedback path frequency response
$G(\omega)$	total sound processor frequency response
$M_{\mathrm{BTE}}(\omega)$	BTE microphone frequency response
$\hat{\mathbf{f}}(t)$	time-domain parameter vector associated with the estimated feedback path model at time t

Chapter 2 notation

n	discrete time index
ω	angular frequency

$A(\omega)$	actuator frequency response
$F(\omega)$	feedback path frequency response
$G(\omega)$	sound processor frequency response
$M(\omega)$	microphone frequency response
$X_{\mathrm{A}}(\omega)$	actuator output signal's frequency spectrum
$R(\omega)$	source signal's frequency spectrum
$y_{\mathrm{M}}[n]$	microphone signal
$r[n]$	source signal
$y[n]$	feedback signal
$x[n]$	actuator input signal
$x_{\mathrm{A}}[n]$	actuator output signal
$x_{\mathrm{ESS}}[n]$	exponential sine sweep signal
$f[n]$	feedback path impulse response
$m[n]$	microphone impulse response
$a[n]$	actuator impulse response
$a_1[n], \ldots, a_K[n]$	Volterra kernels
$h_1[n], \ldots, h_K[n]$	modified Volterra kernels
$\hat{a}_1^{-1}[n]$	estimate of the inverse first order Volterra kernel
$\hat{h}[n]$	deconvolved total response
$\hat{h}_1[n]$	estimate of the first order modified Volterra kernel, i. e. deconvolved impulse response
$\hat{h}_2[n], \ldots, \hat{h}_K[n]$	estimates of the higher-order modified Volterra kernels
$[n_{\mathrm{s}}, n_{\mathrm{e}}]$	noise-only portion of the deconvolved total response
γ	noise threshold
D	distortion measure

Chapter 3 notation

Chapter 3 cannot be shared publicly due to confidentiality obligations.

Chapter 4 notation

l	discrete frequency index
k	discrete frame index
t	discrete time index
$d[t, \hat{\mathbf{f}}(t)]$	time-domain error signal
$e(t)$	time-domain white noise excitation signal
$\varepsilon[t, \hat{\mathbf{a}}(t), \hat{\mathbf{f}}(t)]$**,** $\varepsilon(t)$	time-domain prefiltered error signal
$u(t)$	time-domain loudspeaker signal
$u_a[t, \hat{\mathbf{a}}(t)]$**,** $u_a(t)$	time-domain prefiltered loudspeaker signal
$v(t)$	time-domain source signal
$x(t)$	time-domain feedback signal
$y(t)$	time-domain microphone signal
$\hat{y}[t\|\hat{\mathbf{f}}(t)]$	time-domain predicted feedback signal
$y_a[t, \hat{\mathbf{a}}(t)]$**,** $y_a(t)$	time-domain prefiltered microphone signal
$F(q,t)$	true feedback path
$\hat{F}(q,t)$	estimated feedback path model
$G(q,t)$	electroacoustic forward path model
$H(q,t)$	true source signal generation system model
$A(q,t)$	true inverse source signal model
$\hat{A}(q,t)$	estimated inverse source signal model
$n_{\hat{F}}$	estimated feedback path length, true feedback path length
n_H	number of autoregressive coefficients minus one of $\hat{\mathbf{a}}(k)$
α	regularization factor

μ_{fix}	fixed part of the frequency-domain adaptive stepsize
K_1, K_2	gain values used in the simulations
L	frame shift
M	frame size of the loudspeaker signal
R	frame size of the microphone and error signals
$\hat{\mathbf{a}}(k)$	time-domain estimated prefilter at frame k
$\mathbf{d}(k)$	time-domain error signal at frame k
$\mathbf{e}(k)$	time-domain source excitation signal at frame k
$\hat{\mathbf{f}}(k)$	time-domain estimated feedback path model at frame k
$\mathbf{u}(k)$	time-domain loudspeaker signal at frame k
$\mathbf{u}_a(k)$	time-domain prefiltered [through $\hat{A}(q,t)$] loudspeaker signal at frame k
$\mathbf{y}_a(k)$	time-domain prefiltered [through $\hat{A}(q,t)$] microphone signal at frame k
$\hat{\mathbf{y}}_a(k)$	time-domain prefiltered [through $\hat{A}(q,t)$] predicted feedback signal at frame k
$\boldsymbol{\varepsilon}(k)$	time-domain prediction error signal at frame k
$\mathbf{0}_{M\times M}$	null matrix of dimension $M \times M$
$\mathbf{0}_{M\times 1}$	null vector of dimension $M \times 1$
$\mathbf{F}_M$	unitary DFT matrix of size $M \times M$
$\mathbf{G}_{n_H}$	linearization square matrix, i. e. $\mathbf{G}_{n_H} = \mathbf{F}_M \, \mathbf{Q}_{n_H} \mathbf{Q}_{n_H}^{\mathrm{T}} \mathbf{F}_M^{-1}$ in the FD-PEMAF, and $\mathbf{G}_{n_H} \approx L/M \cdot \mathbf{I}_M$ in the uFD-PEMAF
$\mathbf{I}_M$	identity matrix of dimension $M \times M$
$\mathbf{Q}$	projection matrix used in the PEM-FDAF, i. e. $\mathbf{Q} = [\mathbf{0}_R \quad \mathbf{I}_R]^{\mathrm{T}}$
$\mathbf{Q}_{n_H}$	projection matrix used in the FD-PEMAF, i. e. $\mathbf{Q}_{n_H} = [\mathbf{0}_{L\times R+n_H} \quad \mathbf{I}_L]^{\mathrm{T}}$
$\hat{\mathbf{A}}(k)$	frequency-domain estimated prefilter at frame k
$\mathbf{D}(k)$	frequency-domain error signal at frame k
$\hat{\mathbf{F}}(k)$	frequency-domain estimated feedback path model at frame k
$\Delta\hat{\mathbf{F}}(k)$	frequency-domain estimated update of the adaptive feedback estimate at frame k
$\mathbf{U}(k)$	frequency-domain loudspeaker signal at frame k
$\mathbf{U}_a(k)$	frequency-domain prefiltered [through $\hat{A}(q,k)$] loudspeaker signal at frame k

$\mathbf{Y}_a(k)$	frequency-domain prefiltered [through $\hat{A}(q,k)$] microphone signal at frame k
$\boldsymbol{\mathcal{E}}(k)$	frequency-domain prediction error signal at frame k
$\mu_p(k,l)$	frequency-domain adaptive stepsize for the PEM-FDAF algorithm at frame k and frequency l
$\boldsymbol{\mu}(k)$	frequency-domain adaptive stepsize at frame k
$\boldsymbol{\mu}_a(k)$	frequency-domain prefiltered adaptive stepsize at frame k

Chapter 5 notation

l	discrete frequency index
n	summation index, identifiability conditions section
κ	discrete frame index
t	discrete time index
$d[t,\hat{\mathbf{f}}(t)]$	time-domain error signal
$e(t)$	time-domain white noise source excitation signal
$u(t)$	time-domain loudspeaker signal
$u_{\hat{\mathbf{j}}}[t,\hat{\mathbf{j}}(t)]$	time-domain prefiltered loudspeaker signal
$v(t)$	time-domain source signal
$x(t)$	time-domain feedback signal
$y(t)$	time-domain microphone signal
$\bar{y}[t\|\hat{\mathbf{f}}(t)]$	time-domain predicted feedback signal
$y_{\hat{\mathbf{j}}}[t,\hat{\mathbf{j}}(t)]$	time-domain prefiltered microphone signal
$\varepsilon[t,\boldsymbol{\theta}(t)]$	time-domain prediction error signal parameterized in $\boldsymbol{\theta}(t)$
$\varepsilon[t,\boldsymbol{\xi}(t)]$	time-domain prediction error signal parameterized in $\boldsymbol{\xi}(t)$
$A(q,t)$	model introduced to linearize the first term of the prediction error
$\bar{A}(q,t)$	one-tap-compensated $A(q,t)$
$B(q,t)$	model introduced to linearize the second term of the prediction error
$\bar{B}(q,t)$	delay-compensated $B(q,t)$
$F_{\mathrm{t}}(q,t)$	true feedback path

$F(q,t)$	feedback path model
$\hat{F}(q,t)$	estimated feedback path model
$\bar{F}_\mathrm{t}(q,t)$	delay-compensated true feedback path
$G(q,t)$	electroacoustic forward path model
$\bar{G}(q,t)$	delay-compensated $G(q,t)$
$H_\mathrm{t}(q,t)$	true source signal generation system
$H(q,t)$	source signal generation system model
$\hat{H}(q,t)$	estimated source signal generation system model
$J_\mathrm{t}(q,t)$	model satisfying the equation $J_\mathrm{t}(q,t)H_\mathrm{t}(q,t)=1$
$J(q,t)$	model satisfying the equation $J(q,t)H(q,t)=1$
$Z(q,t)$	model introduced to express the prediction error signal as a function of $v(t)$
$\mathbf{a}(t)$	time-domain parameter vector associated with the model $A(q,t)$ at time t
$\mathbf{b}(t)$	time-domain parameter vector associated with the model $B(q,t)$ at time t
$\mathbf{f}_\mathrm{t}(t)$	time-domain parameter vector associated with the true feedback path at time t
$\mathbf{f}(t)$	time-domain parameter vector associated with the feedback path model at time t
$\hat{\mathbf{f}}(t)$	time-domain parameter vector associated with the estimated feedback path model at time t
$\mathbf{j}_\mathrm{t}(t)$	time-domain parameter vector associated with the model $J_\mathrm{t}(q,t)$ at time t
$\mathbf{j}(t)$	time-domain parameter vector associated with the model $J(q,t)$ at time t
$\boldsymbol{\theta}(t)$	time-domain augmented parameter vector containing $\mathbf{f}(t)$ and $\mathbf{j}(t)$ at time t, i. e. $\boldsymbol{\theta}(t)=[\mathbf{f}^\mathrm{T}(t) \quad \mathbf{j}^\mathrm{T}(t)]^\mathrm{T}$
$\boldsymbol{\xi}(t)$	time-domain augmented parameter vector containing $\mathbf{a}(t)$ and $\mathbf{b}(t)$ at time t, i. e. $\boldsymbol{\xi}(t)=[\mathbf{a}^\mathrm{T}(t) \quad \mathbf{b}^\mathrm{T}(t)]^\mathrm{T}$
d_F	cancellation path delay
d_G	forward path delay
n_A	number of elements minus one of the parameter vector $\mathbf{a}(t)$
n_B	number of elements minus one of the parameter vector $\mathbf{b}(t)$

n_F	number of elements of the parameter vector $\mathbf{f}(t)$, i. e. feedback path model order plus one
$n_{\hat{F}}$	number of elements of the parameter vector $\hat{\mathbf{f}}(t)$, i. e. estimated feedback path model order plus one
n_J	number of elements minus one of the parameter vector $\mathbf{j}(t)$, i. e. model order of $J(q,t)$
n_θ	number of elements of the parameter vector $\boldsymbol{\theta}(t)$
α_{t}	true transition factor
α	transition factor
γ	overlapping factor
δ	regularization factor
μ_{fix}	fixed part of the frequency-domain adaptive stepsize
D	FFT/IFFT operation complexity
K_{MSG}	maximum stable gain of the system when no feedback canceller is included
K_1	gain value used in the simulations, corresponding to $K_{\mathrm{MSG}} - 3\,\mathrm{dB}$
L	partitioned block size
M	frame size
P	partitioned block number per frame
R	frame shift
V	partitioned block shift
$\mathbf{d}[\kappa\|\hat{\mathbf{f}}(\kappa)]$	time-domain error signal at frame κ
$\mathbf{e}(\kappa)$	time-domain white noise source excitation signal at frame κ
$\mathbf{f}_{\mathrm{t}}(\kappa)$	time-domain true feedback path at frame κ
$\hat{\mathbf{f}}(\kappa)$	time-domain estimated feedback path model at frame κ
$\mathbf{f}_{\mathrm{r}}(\kappa)$	time-domain estimation error on the estimated feedback path model at frame κ
$\mathbf{f}_{\mathrm{t},p}(\kappa)$	time-domain partitioned-block true feedback path at frame κ for block p
$\mathbf{f}_{\mathrm{t}}(0)$	time-domain initial true feedback path at frame 0
$\mathbf{j}(\kappa)$	time-domain parameter vector associated with the model $J(q,\kappa)$ [estimated prefilter] at frame κ
$\hat{\mathbf{j}}(\kappa)$	time-domain parameter vector associated with the estimated model $\hat{J}(q,\kappa)$ [estimated prefilter] at frame κ
$\mathbf{n}_{\mathrm{t}}(\kappa)$	time-domain true process noise of the feedback path model at frame κ

$\mathbf{n}(\kappa)$	time-domain model process noise of the feedback path model at frame κ
$\mathbf{u}(\kappa)$	time-domain loudspeaker signal at frame κ
$\mathbf{u}_{\mathrm{J_t}}(\kappa)$	time-domain prefiltered [through $J_\mathrm{t}(q,\kappa)$] loudspeaker signal at frame κ
$\mathbf{u}_{\hat{\mathrm{J}}}(\kappa)$	time-domain prefiltered [through $\hat{J}(q,\kappa)$] loudspeaker signal at frame κ
$\mathbf{u}_p(\kappa)$	time-domain partitioned-block loudspeaker signal at frame κ for block p
$\mathbf{u}_{\mathrm{J_t},p}(\kappa)$	time-domain prefiltered [through $J_\mathrm{t}(q,\kappa)$] partitioned-block loudspeaker signal at frame κ for block p
$\mathbf{u}_{\hat{\mathrm{J}},p}(\kappa)$	time-domain prefiltered [through $\hat{J}(q,\kappa)$] partitioned-block loudspeaker signal at frame κ for block p
$\mathbf{y}_{\mathrm{J_t}}(\kappa)$	time-domain prefiltered [through $J_\mathrm{t}(q,\kappa)$] microphone signal at frame κ
$\mathbf{y}_{\hat{\mathrm{J}}}(\kappa)$	time-domain prefiltered [through $\hat{J}(q,\kappa)$] microphone signal at frame κ
$\bar{\mathbf{y}}_{\hat{\mathrm{J}}}(\kappa)$	time-domain prefiltered [through $\hat{J}(q,\kappa)$] predicted feedback signal at frame κ
$\boldsymbol{\zeta}(\kappa)$	time-domain matrix used to compensate for power variations in the white noise source excitation signal $\mathbf{e}(\kappa)$ at frame κ
$\boldsymbol{\varepsilon}[\kappa,\boldsymbol{\theta}_\mathrm{t}(\kappa)]$	time-domain prediction error signal parameterized in $\boldsymbol{\theta}_\mathrm{t}(\kappa)$ at frame κ
$\boldsymbol{\varepsilon}[\kappa,\boldsymbol{\theta}(\kappa)]$	time-domain prediction error signal parameterized in $\boldsymbol{\theta}(\kappa)$ at frame κ
$\boldsymbol{\varepsilon}[\kappa,\hat{\boldsymbol{\theta}}(\kappa)]$	time-domain prediction error signal parameterized in $\hat{\boldsymbol{\theta}}(\kappa)$ at frame κ
$\boldsymbol{\theta}_\mathrm{t}(\kappa)$	time-domain augmented parameter vector containing $\mathbf{f}_\mathrm{t}(\kappa)$ and $\mathbf{j}_\mathrm{t}(\kappa)$ at frame κ, i. e. $\boldsymbol{\theta}_\mathrm{t}(\kappa) = [\mathbf{f}_\mathrm{t}^\mathrm{T}(\kappa) \quad \mathbf{j}_\mathrm{t}^\mathrm{T}(\kappa)]^\mathrm{T}$
$\boldsymbol{\theta}(\kappa)$	time-domain augmented parameter vector containing $\mathbf{f}(\kappa)$ and $\mathbf{j}(\kappa)$ at frame κ, i. e. $\boldsymbol{\theta}(\kappa) = [\mathbf{f}^\mathrm{T}(\kappa) \quad \mathbf{j}^\mathrm{T}(\kappa)]^\mathrm{T}$
$\hat{\boldsymbol{\theta}}(\kappa)$	time-domain augmented parameter vector containing $\hat{\mathbf{f}}(\kappa)$ and $\hat{\mathbf{j}}(\kappa)$ at frame κ, i. e. $\hat{\boldsymbol{\theta}}(\kappa) = [\hat{\mathbf{f}}^\mathrm{T}(\kappa) \quad \hat{\mathbf{j}}^\mathrm{T}(\kappa)]^\mathrm{T}$
$\boldsymbol{\Lambda}_\mathrm{t}(\kappa)$	time-domain covariance matrix of the process noise $\mathbf{n}_\mathrm{t}(\kappa)$ at frame κ
$\boldsymbol{\mu}_{\mathbf{f}_\mathrm{t}(0)}$	time-domain mean vector of the initial true feedback path $\mathbf{f}_\mathrm{t}(0)$ at frame 0
$\boldsymbol{\Pi}_\mathrm{t}(0)$	time-domain covariance matrix of the initial true feedback path $\mathbf{f}_\mathrm{t}(0)$ at frame 0
$\boldsymbol{\Sigma}_\mathrm{t}(\kappa)$	time-domain covariance matrix of the white noise source excitation signal $\mathbf{e}_\mathrm{t}(\kappa)$ at frame κ
$\mathbf{0}_{M\times M}$	null matrix of dimension $M \times M$
$\mathbf{0}_{M\times 1}$	null vector of dimension $M \times 1$
$\mathbf{1}_{M\times 1}$	vector of ones of dimension $M \times 1$

$\mathbf{F}_M$ unitary DFT matrix of size $M \times M$

$\mathbf{G}^{01}_{R\times M}$ Linearization rectangular matrix, i. e. $\mathbf{G}^{01}_{R\times M} = \mathbf{F}_R \, \mathbf{W}^{01}_{R\times M} \, \mathbf{F}^{-1}_M$

$\mathbf{G}^{10}_{M\times R}$ Linearization rectangular matrix, i. e. $\mathbf{G}^{10}_{M\times R} = \mathbf{F}_M \, \mathbf{W}^{10}_{M\times R} \, \mathbf{F}^{-1}_R$

$\mathbf{G}^{01}_{M\times M}$ Linearization square matrix, i. e. $\mathbf{G}^{01}_{M\times M} = (\mathbf{G}^{01}_{R\times M})^{\mathrm{H}} \mathbf{G}^{01}_{R\times M} = \mathbf{F}_M \, \mathbf{W}^{01}_{M\times M} \, \mathbf{F}^{-1}_M$

$\mathbf{G}^{10}_{M\times M}$ Linearization square matrix, i. e. $\mathbf{G}^{10}_{M\times M} = \mathbf{G}^{10}_{M\times R} (\mathbf{G}^{10}_{M\times R})^{\mathrm{H}} = \mathbf{F}_M \, \mathbf{W}^{10}_{M\times M} \, \mathbf{F}^{-1}_M$

$\mathbf{I}_M$ identity matrix of dimension $M \times M$

$\mathbf{W}^{01}_{R\times M}$ Constraint rectangular matrix expanding to fat matrix, i. e. $\mathbf{W}^{01}_{R\times M} = \begin{bmatrix} \mathbf{0}_{R\times R} & \mathbf{I}_R \end{bmatrix}$

$\mathbf{W}^{10}_{M\times R}$ Constraint rectangular matrix expanding to tall matrix, i. e. $\mathbf{W}^{10}_{M\times R} = \begin{bmatrix} \mathbf{I}_R \\ \mathbf{0}_{R\times R} \end{bmatrix}$

$\mathbf{W}^{01}_{M\times M}$ Constraint rectangular matrix selecting lower square corner, i. e. $\mathbf{W}^{01}_{M\times M} = \begin{bmatrix} \mathbf{0}_{R\times R} & \mathbf{0}_{R\times R} \\ \mathbf{0}_{R\times R} & \mathbf{I}_R \end{bmatrix}$

$\mathbf{W}^{10}_{M\times M}$ Constraint rectangular matrix selecting upper square corner, i. e. $\mathbf{W}^{10}_{M\times M} = \begin{bmatrix} \mathbf{I}_R & \mathbf{0}_{R\times R} \\ \mathbf{0}_{R\times R} & \mathbf{0}_{R\times R} \end{bmatrix}$

$\mathbf{C}_{\mathrm{J_t}}(\kappa)$ frequency-domain prefiltered [through $J_{\mathrm{t}}(q,\kappa)$] loudspeaker signal including the linearization operation at frame κ

$\mathbf{C}_{\hat{\mathrm{J}}}(\kappa)$ frequency-domain prefiltered [through $\hat{J}(q,\kappa)$] loudspeaker signal including the linearization operation at frame κ

$\mathbf{C}_{\mathrm{J_t},p}(\kappa)$ frequency-domain prefiltered [through $J_{\mathrm{t}}(q,\kappa)$] partitioned-block loudspeaker signal including the linearization operation at frame κ for block p

$\mathbf{C}_{\hat{\mathrm{J}},p}(\kappa)$ frequency-domain prefiltered [through $\hat{J}(q,\kappa)$] partitioned-block loudspeaker signal including the linearization operation at frame κ for block p

$\mathbf{E}(\kappa)$ frequency-domain white noise source excitation signal at frame κ

$\mathbf{F}_{\mathrm{t}}(\kappa)$ frequency-domain true feedback path at frame κ

$\mathbf{F}(\kappa)$ frequency-domain feedback path model at frame κ

$\hat{\mathbf{F}}(\kappa)$ frequency-domain estimated feedback path model at frame κ

$\mathbf{F}_{\mathrm{r}}(\kappa)$ frequency-domain estimation error on the estimated feedback path model at frame κ

$\hat{\mathbf{F}}^{+}(\kappa)$	frequency-domain a-posteriori estimated feedback path model at frame κ
$\mathbf{F}_{\mathrm{t},p}(\kappa)$	frequency-domain partitioned-block true feedback path at frame κ for block p
$\hat{\mathbf{F}}_{p}(\kappa)$	frequency-domain partitioned-block estimated feedback path model at frame κ for block p
$\hat{\mathbf{F}}_{p}^{+}(\kappa)$	frequency-domain partitioned-block a-posteriori estimated feedback path model at frame κ for block p
$\mathbf{F}(0)$	frequency-domain feedback path model at frame 0
$F_{\mathrm{r}}(\kappa, l)$	frequency-domain estimation error on the estimated feedback path at frame κ and frequency l
$\mathbf{J}(\kappa)$	frequency-domain parameter vector associated with the model $J(q, \kappa)$ at frame κ
$\mathbf{K}(\kappa)$	frequency-domain Kalman gain at frame κ
$\mathbf{K}_{p}(\kappa)$	frequency-domain partitioned-block Kalman gain at frame κ for block p
$\mathbf{L}_{\mathrm{t}}(\kappa)$	frequency-domain covariance matrix of the true process noise $\mathbf{N}_{\mathrm{t}}(\kappa)$ at frame κ
$\mathbf{N}_{\mathrm{t}}(\kappa)$	frequency-domain true process noise of the feedback path model at frame κ
$\mathbf{N}(\kappa)$	frequency-domain process noise of the feedback path model at frame κ
$\mathbf{N}_{\mathrm{t},p}(\kappa)$	frequency-domain true partitioned-block process noise of the feedback path model at frame κ for block p
$\mathbf{P}_{\mathrm{t}}(0)$	frequency-domain covariance matrix of the true feedback path $\mathbf{F}_{\mathrm{t}}(0)$ at frame 0
$\mathcal{P}(\kappa)$	set of frequencies satisfying the phase condition of the Nyquist stability criterion at frame κ
$\mathbf{P}(\kappa)$	frequency-domain error covariance matrix at frame κ
$\mathbf{P}(0)$	frequency-domain error covariance matrix at frame 0, initialization value
$\mathbf{P}^{+}(\kappa)$	frequency-domain a-posteriori error covariance matrix at frame κ
$\mathbf{P}_{p}(\kappa)$	frequency-domain partitioned-block error covariance matrix at frame κ for block p
$\mathbf{P}_{p}^{+}(\kappa)$	frequency-domain partitioned-block a-posteriori error covariance matrix at frame κ for block p
$\mathbf{\Psi}_{\mathrm{EE}}(\kappa)$	frequency-domain covariance matrix of $\mathbf{E}(\kappa)$ at frame κ
$\mathbf{\Psi}_{\hat{\mathrm{E}}\hat{\mathrm{E}}}(\kappa)$	frequency-domain estimate of the covariance matrix $\mathbf{\Psi}_{\mathrm{EE}}(\kappa)$ at frame κ
$\mathbf{\Phi}_{\hat{\mathrm{E}}\hat{\mathrm{E}}}(\kappa)$	frequency-domain power spectral density estimate of $\mathbf{E}(\kappa)$ at frame κ
$\mathbf{\Psi}_{\mathrm{N_t N_t}}(\kappa)$	frequency-domain covariance matrix of $\mathbf{N}_{\mathrm{t}}(\kappa)$ at frame κ

$\mathbf{\Psi}_{\mathbf{N}_\mathrm{t}\mathbf{N}_\mathrm{t},\mathrm{p}}(\kappa)$	frequency-domain partitioned-block covariance matrix of $\mathbf{N}_\mathrm{t}(\kappa)$ at frame κ for block p
$\mathbf{\Psi}_{\hat{\mathbf{N}}\hat{\mathbf{N}}}(\kappa)$	frequency-domain estimate of the covariance matrix $\mathbf{\Psi}_{\mathbf{N}_\mathrm{t}\mathbf{N}_\mathrm{t}}(\kappa)$ at frame κ
$\mathbf{\Psi}_{\hat{\mathbf{N}}\hat{\mathbf{N}},\mathrm{p}}(\kappa)$	frequency-domain estimate of the partitioned-block covariance matrix $\mathbf{\Psi}_{\mathbf{N}_\mathrm{t}\mathbf{N}_\mathrm{t},\mathrm{p}}(\kappa)$ at frame κ for block p
$\mathbf{\Phi}_{\hat{\mathbf{N}}\hat{\mathbf{N}}}(\kappa)$	frequency-domain power spectral density estimate of $\mathbf{N}_\mathrm{t}(\kappa)$ at frame κ
$\mathbf{\Phi}_{\hat{\mathbf{N}}\hat{\mathbf{N}},\mathrm{p}}(\kappa)$	frequency-domain partitioned-block power spectral density estimate of $\mathbf{N}_{\mathrm{t},p}(\kappa)$ at frame κ for block p
$\mathbf{\Phi}_{\hat{\mathcal{E}}\hat{\mathcal{E}}}(\kappa)$	frequency-domain power spectral density estimate of $\boldsymbol{\mathcal{E}}[\kappa, \hat{\mathbf{\Theta}}(\kappa)]$ at frame κ
$\mathbf{S}_\mathrm{t}(\kappa)$	frequency-domain covariance matrix of the true white noise excitation signal $\mathbf{E}(\kappa)$ at frame κ
$\mathbf{U}_{\mathrm{J}_\mathrm{t}}(\kappa)$	frequency-domain prefiltered [through $J_\mathrm{t}(q,\kappa)$] loudspeaker signal at frame κ
$\mathbf{U}_{\hat{\mathrm{J}}}(\kappa)$	frequency-domain prefiltered [through $\hat{J}(q,\kappa)$] loudspeaker signal at frame κ
$\mathbf{U}_{\mathrm{J}_\mathrm{t},p}(\kappa)$	frequency-domain prefiltered [through $J_\mathrm{t}(q,\kappa)$] partitioned-block loudspeaker signal at frame κ for block p
$\mathbf{U}_{\hat{\mathrm{J}},p}(\kappa)$	frequency-domain prefiltered [through $\hat{J}(q,\kappa)$] partitioned-block loudspeaker signal at frame κ for block p
$\mathbf{Y}_{\mathrm{J}_\mathrm{t}}(\kappa)$	frequency-domain prefiltered [through $J_\mathrm{t}(q,\kappa)$] microphone signal at frame κ
$\mathbf{Y}_{\hat{\mathrm{J}}}(\kappa)$	frequency-domain prefiltered [through $\hat{J}(q,\kappa)$] microphone signal at frame κ
$\boldsymbol{\mathcal{E}}[\kappa, \hat{\mathbf{\Theta}}(\kappa)]$	frequency-domain prediction error signal at frame κ parameterized in $\hat{\mathbf{\Theta}}(\kappa)$
$\mathbf{\Theta}(\kappa)$	frequency-domain augmented parameter vector containing $\mathbf{F}(\kappa)$ and $\mathbf{J}(\kappa)$ at frame κ, i. e. $\mathbf{\Theta}(\kappa) = [\mathbf{F}^\mathrm{T}(\kappa) \quad \mathbf{J}^\mathrm{T}(\kappa)]^\mathrm{T}$
$\boldsymbol{\mu}_{\mathbf{F}_\mathrm{t}(0)}$	mean vector of the true feedback path $\mathbf{F}_\mathrm{t}(0)$ at frame 0
$\mu_p(\kappa, l)$	frequency-domain adaptive stepsize for the PEM-PBFDAF algorithm at frame κ and frequency l

Chapter 6 notation

t	discrete time index

f	discrete frequency index
$a(t)$	time-domain anechoic source signal
$d[t, \hat{\mathbf{f}}(t)]$	time-domain error signal
$n(t)$	time-domain additive noise corrupting the source signal
$u(t)$	time-domain loudspeaker signal
$v(t)$	time-domain source signal
$x(t)$	time-domain feedback signal
$y(t)$	time-domain microphone signal
$\bar{y}[t\|\hat{\mathbf{f}}(t)]$	time-domain predicted feedback signal
$F_{\mathrm{t}}(q,t)$	true feedback path
$\hat{F}(q,t)$	estimated feedback path model
$G(q,t)$	electroacoustic forward path
$R(q,t)$	room model filtering the anechoic source to obtain the source signal
$\hat{\mathbf{f}}(t)$	time-domain parameter vector associated with the estimated feedback path model at time t
d_G	forward path delay
$n_{\hat{F}}$	number of elements of the parameter vector $\hat{\mathbf{f}}(t)$, i. e. estimated feedback path model order plus one
α	transition factor of the PEM-FDKF
δ	regularization factor of the PEM-FDAF
μ_{BNLMS}	fixed part of the time-domain adaptive stepsize of the BNLMS
μ_{FDAF}	fixed part of the frequency-domain adaptive stepsize of the PEM-FDAF
μ_{anch}	fixed part of the time-domain adaptive stepsize of the BNLMS used as anchor
$\mathbf{P}(0)$	frequency-domain error covariance matrix at frame 0 of the PEM-FDKF
K_{MSG}	maximum stable gain of the system when no feedback canceller is included
K_1, K_2, K_3	gain values used in the simulations
M	frame size
P	frame length of the prefilter estimation window
R	frame shift

T_s	sampling time
p	p-value
p_A	p-values from the mixed-design ANOVA with Huynh-Feldt correction
p^{o}_{WJ}	p-values from the omnibus Welch-James's test with approximate degrees of freedom
p_T	p-values from the post-hoc Tukey test
p^{p}_{WJ}	p-values from the post-hoc Welch-James's test with approximate degrees of freedom with Hochberg correction
η^2_G	generalized eta-squared effect size
d	Cohen's d effect size
$SD(t)$	frequency-weighted log-spectral signal distortion
w_{ERB}	ERB scale weighting
$P_r(f,t)$	PSD of the reference signals
$P_u(f,t)$	PSD of the processed signals
O_{sig}	sigmoidal-mapped objective scores
O	objective scores
S	subjective scores
α_1, α_2	fitting parameters of the sigmoidal mapping
$\hat{\alpha}_1$, $\hat{\alpha}_2$	estimated fitting parameters of the sigmoidal mapping
r	Pearson correlation coefficient
r_{spear}	Spearman correlation coefficient
r_{sig}	sigmoid-mapped Pearson correlation coefficient
RMSE	root-mean squared error
ε**-RMSE**	epsilon-insensitive root-mean squared error
T_{ij}	value of the statistic used to determine the significance of the ε-RMSE

Contents

Preface i

Abstract v

Korte Inhoud vii

Glossary xi

Contents xxix

List of Figures xxxv

List of Tables xxxix

1 Introduction 1

1.1 Normal hearing and hearing impairment 2

1.1.1 The auditory system 2

1.1.2 Hearing loss 3

1.2 Hearing devices 4

1.3 Feedback in hearing devices 8

1.3.1 Feedback in hearing aids 10

1.3.2 Feedback in implantable hearing devices 12

1.4 Feedback control in hearing devices 13
1.4.1 Adaptive feedback cancellation algorithms 15
1.4.2 Performance assessment of AFC algorithms 16
1.5 Overview of the thesis 18
1.5.1 General overview 18
1.5.2 Thesis outline 19

Part I Feedback Characterization 21

2 The Codacs direct acoustic cochlear implant 23

2.1 Introduction 25
2.2 Materials and methods 28
2.2.1 Codacs actuator transfer function 28
2.2.2 Feedback system description 28
2.2.3 Feedback path estimation 30
2.2.4 Distortion measure 32
2.2.5 Cadaver heads measurements 34
2.2.6 Calibration measurement 36
2.2.7 Data presentation 36
2.3 Results 37
2.3.1 Cadaver heads measurements 37
2.3.2 Application of the microphones calibration measurements 41
2.3.3 Nonlinearity of the system 43
2.4 Discussion 43
2.4.1 Actuator input precompensation 43
2.4.2 Feedback path measurements and specimen-dependent variabilities 44

2.4.3 Effects of an incorrect implantation and nonlinear behavior 45
2.4.4 Difference between structure-borne mechanical and acoustic feedback 47
2.5 Conclusions 48

3 The early prototype of a bone conduction implant concept 49

Part II Adaptive Feedback Cancellation Algorithms 51

4 The frequency-domain PEM-based adaptive filter 53
4.1 Introduction 55
4.2 PEM-based frequency-domain adaptive filter 57
4.3 Frequency-domain PEM-based adaptive filter 59
4.4 Complexity 62
4.5 Simulation Results 63
4.6 Conclusions 65

5 The PEM-based frequency-domain Kalman filter 67
5.1 Introduction 69
5.2 Prediction error method identification 72
5.3 PEM-Based Frequency Domain Kalman Filter 74
5.4 PEM-FDKF Identifiability conditions 80
5.5 PEM-Based Partitioned-Block Frequency Domain Kalman Filter 86
5.6 Computational Complexity 90
5.7 Simulations 93
5.7.1 Stationary feedback path 94
5.7.2 Non-stationary feedback path 97
5.8 Conclusions 99

Part III Subjective and Objective Sound-Quality Evaluation 101

6 Subjective and Objective Sound-Quality Evaluation of Adaptive Feedback Cancellation Algorithms 103

6.1 Introduction . . . 105
6.2 Materials and Methods . . . 108
6.2.1 Algorithms . . . 108
6.2.2 Signal recordings and preprocessing . . . 109
6.2.3 Speech and music material . . . 111
6.2.4 Signals post-processing . . . 112
6.3 Subjective Evaluation . . . 113
6.3.1 Listening test . . . 113
6.3.2 Statistical analysis . . . 116
6.4 Objective Evaluation . . . 117
6.4.1 HAAQI, HASPI, and HASQI . . . 118
6.4.2 PEAQ . . . 118
6.4.3 PEMO-Q . . . 119
6.4.4 PESQ . . . 119
6.4.5 SD . . . 119
6.4.6 STOI . . . 120
6.5 Performance Criteria . . . 120
6.6 Results and Discussion . . . 121
6.6.1 Subjective-evaluation measures . . . 121
6.6.2 Objective measures and performance criteria . . . 129
6.7 Conclusions . . . 132

7 Conclusion 133

Bibliography **141**

Curriculum Vitae **157**

Publication List **159**

List of Figures

1.1 The auditory system. 3

1.2 Classification of hearing loss treatments, based on the type and the degree of hearing loss. 5

1.3 (a): Simplified scheme depicting a general case of feedback. (b): Simplified scheme depicting a case of acoustic feedback in a HA overlayed with a real HA. 11

1.4 Simplified block diagram of an implantable hearing device operating in closed loop. 13

1.5 General AFC scenario in a HA. 14

2.1 Details of the Cochlear™ Codacs™ DACI. 26

2.2 Vibration velocity to voltage sensitivity as a function of frequency for the two Codacs actuators used throughout the four MSs. . . 27

2.3 Simplified block diagram of an implanted Codacs DACI operating in closed-loop. 29

2.4 Example of a DTR in the time domain from one of the measured BTE microphone signals. 32

2.5 MSs setup. 34

2.6 Recorded microphone signals and corresponding spectrograms from BTE and AB1 microphones obtained with a 0.71 V_{RMS} ESS stimulus in MS4. 37

2.7 DIR and corresponding spectrograms from BTE and AB1 microphone signals obtained with a 0.71 V_{RMS} ESS stimulus in MS1 to MS4. 38

2.8 Top: Correct and incorrect placement of the Codacs actuator (top-left and top-right, respectively). 39

2.9 Magnitude response of the DIR from BTE and AB1 microphone signals for the four tested levels in MS1 to MS4. 40

2.10 Calibrated microphone PSD in dB SPL of the BTE and AB1 microphones signals. 42

2.11 Distortion calculated from BTE and AB1 microphone DTR at different levels in MS1 to MS4. 43

4.1 General AFC scenario. 56

4.2 Block diagram of an AFC including PEM-based decorrelation. 57

4.3 Mis and ASG performance for the three algorithms using the two tested gains. 64

5.1 General AFC scenario. 70

5.2 Complete AFC algorithm with PEM stage. 71

5.3 Per-output-sample computational complexity of the existing and proposed algorithms as a function of the prefilter order n_J and the hop size R. 92

5.4 Top: Magnitude responses of the 10 different AIR used in the simulations. Bottom: Magnitude responses of 20 AIR obtained with (5.23a) every 1 s, starting from AIR1. 95

5.5 Mis and ASG performance over time of PEM-FDAF, PEM-FDKF, PEM-PBFDAF and PEM-PBFDKF using the constant forward path gain K_1, AIR1, and 10 different source signals. . . 96

5.6 Mis and ASG performance over time of PEM-FDAF, PEM-FDKF, PEM-PBFDAF and PEM-PBFDKF using the constant forward path gain K_1 and 10 different AIR. 97

5.7 Mis and ASG performance over time of PEM-FDAF, PEM-FDKF, PEM-PBFDAF and PEM-PBFDKF using the constant forward path gain K_1, a smooth AIR transition. 99

5.8 Mis and ASG performance over time of PEM-FDAF, PEM-FDKF, PEM-PBFDAF and PEM-PBFDKF using the constant forward path gain K_1, an abrupt AIR transition from AIR1 to AIR2, and 10 different source signals. 100

6.1 General AFC scenario. 106

6.2 Feedback scenario. 110

6.3 Training and test interfaces of the listening test. 115

6.4 Box plots of the subjective scores for the five different conditions across all different speech signals. 122

6.5 Box plots of the subjective scores for the five different conditions across all different music signals. 123

6.6 Scatter plots of subjective scores versus objective scores, including the sigmoid mapping calculated for the set of data within each scatter plot. 129

List of Tables

4.1 Per-sample computational complexity of the algorithms. 63

4.2 Mean and maximum SD values for the different algorithms in the two tested scenarios. 65

5.1 Definition of the time-invariant constraints used in the paper. . 75

5.2 Definition of linearization matrices used in the paper. 76

5.3 Per-output-sample computational complexity and memory requirements of the compared algorithms. 91

5.4 Mean and maximum SD values for the different algorithms in the five different simulated scenarios. 99

6.1 Test signals IDs, descriptions and employed gains. 111

6.2 Results of the statistical analyses by means of the mixed-design ANOVAs. 124

6.3 Results of the statistical analyses by means of the Welch-James's tests with approximate degrees of freedom. 125

6.4 Results obtained using the different performance criteria when comparing objective and subjective scores. 131

Chapter 1

Introduction

Hearing loss is a debilitating condition that occurs when some form of impairment affects the auditory system. According to the World Health Organization, in February 2017 more than 360 million people worldwide, 32 million of which are children, have a disabling hearing loss [1]. The latter amounts to roughly 5 % of the world's population. When focusing on people older than 65, the percentage of people with a disabling hearing loss increases to roughly 33 % of the world's elderly population due to the natural aging of the auditory system. Considering both the aging of the world's population and the increased exposure to noise in daily life (e. g. through personal music devices such as smartphones, or at entertainment venues such as bars, nightclubs and sporting events), the aforementioned numbers are expected to grow substantially in the coming years [1].

Given the complex structure of the auditory system, a broad spectrum of problems can affect normal hearing. The standard classification, e. g. by Moore *et al.* [2], discriminates between three types of hearing loss: conductive, sensorineural, and mixed loss. Conductive hearing loss (CHL) occurs when issues in the outer ear or the middle ear prevent a proper sound transmission to the inner ear. When the hearing problem originates from an impairment of the inner ear or higher levels of the auditory pathway, it is referred to as sensorineural hearing loss (SNHL). Finally, mixed hearing loss (MHL) is a combination of CHL and SNHL.

In order to partially restore the functionalities resulting from differing kinds of impairments within the auditory system, different types of hearing devices have been developed, each with their strengths and weaknesses. A common problem occurring in different types of hearing devices is that of *feedback*. Feedback arises

from the coupled interaction between the actuator (e. g. an electroacoustic actuator, i. e. a loudspeaker) and the microphone of a hearing device within a closed loop. This gives rise to sound degradation and various artifacts, and it represents a growing problem, due to the miniaturization of hearing devices.

Conventionally, when talking about *feedback*, either in public address (PA) systems or in conventional hearing aids (HAs), the term is usually preceded by the adjective *acoustic*. However, other feedback components might also play a role, especially in the case of implantable hearing devices. Therefore, in the rest of the thesis, feedback is addressed in a more general sense, including a semantic link to the specific component that might be causing it, such as acoustic, structure-borne mechanical, or electromagnetic.

In this thesis, the main focus is on the problem of feedback in different hearing devices and three specific aspects related to this problem are going to be investigated. Part I consists of chapter 2 and chapter 3 and describes the feedback characterization measurements performed on two novel hearing devices. Part II consists of chapter 4 and chapter 5 and describes two novel adaptive feedback cancellation (AFC) algorithms focusing on lower computational complexity, improved performance, and sound quality. Part III consists of chapter 6 and describes an analysis based on subjective listening tests aiming to evaluate the performance in terms of sound quality of one the new AFC algorithms. Finally, the conclusions of the thesis are drawn in chapter 7.

1.1 Normal hearing and hearing impairment

The first portion of this chapter is organized in two parts: the first briefly describing the basic underlying mechanisms of the human auditory system and the second describing how their improper functioning can lead to hearing impairment.

1.1.1 The auditory system

The human auditory system is usually classified in terms of the outer, middle, and inner ear [2], cf. fig. 1.1. In reference to fig. 1.1, the green, blue, and purple sections highlight the outer, middle, and inner ear, respectively. The outer ear consists of the pinna and the ear canal responsible for focusing sound to the tympanic membrane. The middle ear consists of the ossicle chain, including the malleus, the incus and the stapes, which provides an impedance matching between the tympanic membrane and the cochlea, efficiently propagating

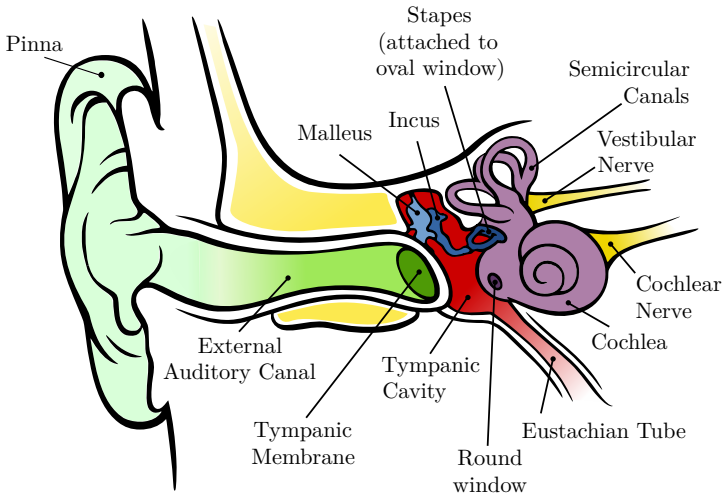


Figure 1.1: The auditory system. Figure By Chittka, L. and Brockmann, A. [CC BY 2.5 (`http://creativecommons.org/licenses/by/2.5`)], via Wikimedia Commons.

airborne mechanical sound waves to the inner ear. The inner ear consists of both the semicircular canals, mainly involved in equilibrioception, and the cochlea. The latter performs the transduction from mechanical signals to electrical (neural) signals through a series of hair cells, organized along the basilar membrane, functioning as mechanoreceptors. Local movements of the fluid filling the cochlea stretch the hair cells at the local position, effectively creating a neural signal that travels through the auditory nerve to higher stages of the brain.

In a simplified interpretation, the operations of the cochlea can be related to those of a frequency analyzer: the basilar membrane has a stiff and wide base, whose hair cells are sensitive to high frequencies, and a loose and narrow apex, whose hair cells are sensitive to low frequencies. This property is usually referred to as the cochlear tonotopic organization.

1.1.2 Hearing loss

As mentioned earlier, there exist different kinds of hearing loss based on the portion of the auditory system that is affected by the impairment. CHL includes

different impairments occurring in the outer and middle ear and can be due to several different causes and pathologies, e. g. cerumen, middle ear infections, perforated eardrum, otosclerosis, benign tumors, congenital malformation of the outer or middle ear [3]. SNHL includes different impairments occurring in the cochlea or in the auditory nerve and can be due to several different causes and pathologies, such as genetic diseases or infections. However, the most common cause of SNHL is presbycusis, or age-related hearing loss, afflicting roughly one third of the world's over-65 population [1].

The negative effect of aging on the auditory system is caused by the damaging of the cochlea hair cells, and such an effect is permanent given that the hair cells do not regenerate. Usually, the hair cells on the base of the basilar membrane, i. e. those sensitive to high frequencies, are the first to be damaged. This is why presbycusis is mostly identified with a high-frequency hearing loss. In an analogous way, prolonged exposure to loud noise can also lead to the damage of the hair cells. This is why there are raising concerns about the improper exposure to loud noise, both in working environments and for recreational means [1].

Based on the hearing loss range, expressed in dB HL, one can identify the degree of hearing loss a patient is suffering from, i. e. mild, moderate, moderately severe, severe, profound, or deaf (cf. fig. 1.2).

1.2 Hearing devices

Due to the different nature of the impairments affecting the auditory system, there exist several types of hearing devices. Figure 1.2 illustrates how different types of hearing devices/therapies can be used to treat kinds of hearing loss, based on the type and on the degree of hearing loss of a patient. The following hearing devices/therapies can be identified: HAs, middle ear implants (MEIs), middle ear surgery (MES), bone-conduction devices (BCDs), direct acoustic cochlear implants (DACIs), and cochlear implants (CIs). Despite the existence of all these types of hearing devices, recent data estimate that the global production of hearing devices meets less than 10 % of the global need worldwide, and less than 3 % of the need in developing countries [1]. The estimated annual global cost due to unaddressed hearing loss amounts to 750 billion international dollars [1].

Conventional HAs represent the most common treatment for mild to severe SNHL. However, they can also be used to treat mild MHL. Their popularity is due to the fact that they address the most widespread hearing impairments, such as age-related hearing loss or exposure hearing loss, and due to their earlier

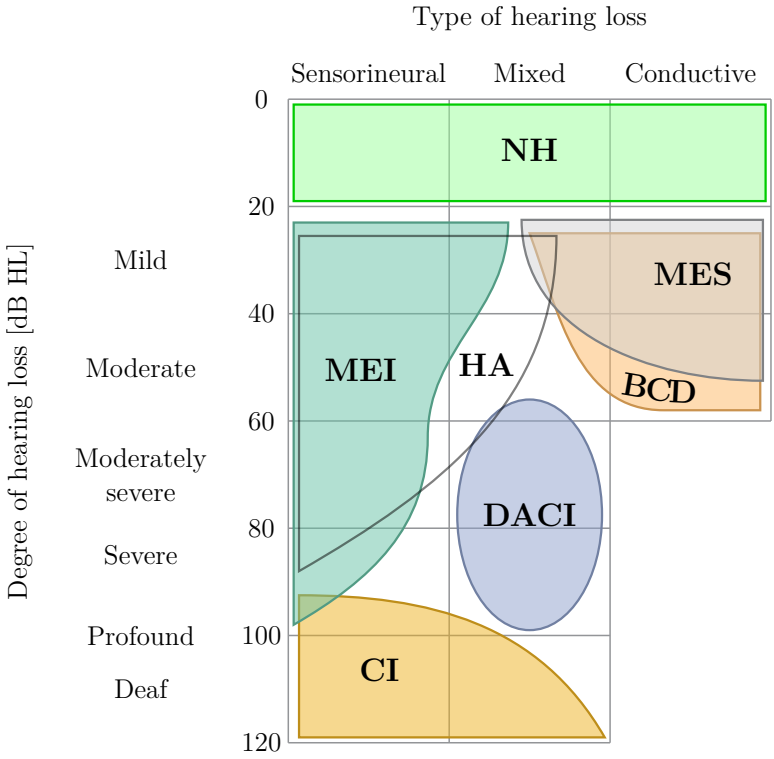


Figure 1.2: Classification of hearing loss treatments, based on the type and the degree of hearing loss: hearing aids (HAs), middle ear implants (MEIs), middle ear surgery (MES), bone-conduction devices (BCDs), direct acoustic cochlear implants (DACIs), and cochlear implants (CIs). The area indicating normal hearing (NH) is also indicated. Adapted from fig. 1 in [4], Copyright © 2011, IEEE.

introduction compared to other hearing devices [5]. Additionally, compared to other types of hearing devices, they are less invasive since they do not require surgery. A standard HA works by means of three main stages: 1) One or more microphones (at the time of writing a two-to-three microphone array is the norm [6]) record the sound scene surrounding the HA recipient; 2) The sound scene recordings are processed through different algorithms embodied in a sound processor; 3) The processed audio signal is played back through a loudspeaker, whose tip is inserted in the patient's ear canal.[1] The sound processor can be of

[1]Even though in HA literature, the *loudspeaker* is often referred to as *receiver*, the word *loudspeaker* will be used throughout this dissertation for the sake of consistency.

different sizes and can be placed in different positions based on the HA type. The different algorithms implemented in the sound processor work to compensate for the impairments of the HA recipient, e. g. by providing noise reduction, automatic gain control, sound scene classification, and nonlinear compression, as well as, in some cases, to reduce problems introduced by the HA itself, e. g. feedback. Given the high demand, many companies manufacture and sell HAs such as Bernafon (subsidiary of the William Demant Holding Group, Smørum, Denmark), GN Resound (Ballerup, Denmark), Oticon (subsidiary of the William Demant Holding Group, Smørum, Denmark), Sivantos (Singapore), Phonak (subsidiary of the Sonova Holding AG, Stäfa, Switzerland), Starkey Hearing Technologies (Eden Praire, MN, USA), and Widex (Lynge, Allerød Municipality, Denmark), among others.

MES is used to treat mild to moderate forms of CHL, through passive middle-ear prostheses. The prosthesis can replace either part, e. g. the stapes [7], or the whole ossicle chain [8].

MEIs, like the Carina© (Cochlear Ltd, Sydney, Australia[2]) [10], the Vibrant Soundbridge® (MED-EL, Innsbruck, Austria) [11, 12], the Esteem® (Enovy, St. Paul, MN, USA) [13, 14], and the MAXUM (Ototronix, Houston, TX, USA[3]) [16, 17], do not rely on a loudspeaker to provide the sound signal to the patient. They, instead, employ an electromechanical actuator mechanically coupled to one of the (healthy) middle-ear ossicles. Such an actuator amplifies the incoming sound signal by amplifying the movements of the bones in the intact ossicle chain through vibrations. Being an alternative to conventional HAs, MEIs are also used to treat moderate to severe SNHL [18], but not CHL or MHL. The sound processor might be externally worn as for a conventional HA, implanted, or placed in the ear canal.

BCDs are devices which, as the name suggests, exploit bone conduction to restore the functionality of the impaired auditory system [19, 20]. They are mainly used to treat CHL even though they might be also used to treat mild MHL. There exist many different kinds of BCDs and a simplified categorization is based on whether they transmit vibrations to the skin (i. e. skin-drive) or to the skull bone directly (i. e. direct-drive) [21]. Regardless of their characteristics, all BCDs, at the time of writing, are characterized by a vibrating part and a sound processor. The vibrating part can be either worn, semi or fully implantable, and transduces a voltage into a force. The sound processor can either be physically attached to the vibrating part or only electrically coupled; it includes one or more microphones and may exploit different algorithms to denoise, compress, and provide feedback control. Through a BCD, both the

[2]The Cochlear Carina is a technological evolution of the Otologics MET [9].
[3]The Ototronix MAXUM was formerly known as SOUNDTEC Direct System [15].

outer and the middle ear are bypassed. The Baha® 3 Power and Baha® 4 (Cochlear Bone Anchored Solutions AB, Mölnlycke, Sweden), and the Ponto (Oticon Medical, Askim, Sweden) bone-anchored hearing aids (BAHAs) [19, 20, 22, 23, 24], as well as the Bonebridge™ (MED-EL, Innsbruck, Austria) [25, 26, 27], and the bone-conduction implant (BCI) [28, 29, 30, 31, 32] are all examples of commercial BCDs. The early prototype of a bone conduction implant concept (BCIC) studied in this thesis is also a BCD.

The DACI, like the Codacs™ (Cochlear Ltd., Sydney, Australia), is a relatively new implantable hearing device that was developed to treat severe to profound MHL [33, 4]. The DACI consists of an implantable electro-mechanical actuator, placed within the mastoid cavity and firmly fixed through a bone plate directly stimulating the cochlea via a piston prosthesis coupled to an artificial actuator [3]. Unlike MEIs, the DACI does not require a healthy ossicles chain (the chain is in fact interrupted during the implantation procedure) as the stimulation is directly applied to the cochlea. A standard behind-the-ear (BTE) sound processor is used, which communicates with the implant through a radio-frequency (RF) link.

Finally, unlike the previous hearing devices, CIs do not rely on mechanical stimuli (neither through vibrations nor through acoustic waves) to restore the hearing impairment of the patient. This is because for patients with a profound hearing loss or with complete deafness there are little to none remaining functional hair cells. Therefore, their hair cells' function cannot be compensated by simply providing a larger amplification. Alternatively, provided that the cochlear nerve fibers are still functional, a CI can be used to partially restore that patient's hearing, by directly stimulating the cochlea at different places through an electrode array. As most of the previously described implants, a CI relies on a sound processor implementing different audio enhancement algorithms. After this preprocessing the audio signal is coded in a frequency-dependent manner and transmitted through an RF link to the aforementioned intracochlear electrode array. Coding algorithms, necessary to provide the proper stimuli at the different electrodes, are especially important for CIs. The four main cochlear implants manufacturers worldwide are Advanced Bionics (subsidiary of the Sonova Holding AG, Stäfa, Switzerland), Cochlear Ltd. (Sydney, Australia), MED-EL (Innsbruck, Austria), and Oticon Medical (subsidiary of the William Demant Holding Group, Smørum, Denmark).

Multiple hearing devices of the aforementioned types can also be used in combination, leading, e. g., to patients fitted with bilateral HAs or bilateral CIs. Alternatively, bilateral hearing can also be achieved bimodally, i. e. fitting the patient with two different hearing devices, e. g. a HA and a CI. At the time of writing, the trend of fitting patients with a combination of hearing devices when necessary is increasing and, for instance, bilateral CIs are already

reimbursed in a number of countries worldwide. This is beneficial since it leads to bilateral hearing and may, possibly, lead to a binaural improvement, such as an improved performance in spatial localization tasks and binaural noise reduction [34, 35, 36, 6]. However, other studies have pointed out that the presence of several variables involved in the success of a bilateral/bimodal configuration, in some cases may also even lead to worse performance [37, 38, 39, 40]. An important factor that seems to be often related to the success of a specific bilateral/bimodal configuration is the combined tuning of both the hearing devices, e. g. through program or volume control settings and gain-control parameters, and even through audio signals sharing between hearing devices [41]. Some of these tuning tasks are performed during the hearing device fitting, while others are inherently adaptive and are hence algorithmically controlled throughout the usage of the hearing devices. However, this is not always an easy task due to restrictions in terms of computational complexity of the binaural algorithms, communication delays between the hearing devices, and the difficulties of including high-rate wireless links between different hearing devices [6]. These restrictions are still, at the time of writing, a limiting factor in the number of fully binaural configurations available on the market. Nevertheless, with the advancing technology in the hearing devices, both from a software and a hardware point of view, exploiting binaural processing may become more widespread in the near future.

1.3 Feedback in hearing devices

Nowadays hearing devices employ a collection of different algorithms in an attempt to compensate for the patient's hearing impairment. The main signal processing techniques used for this purpose include, among others, (multi-microphone) noise reduction, dynamic range compression, spectral contrast enhancement, acoustic scene, and sound classification [42, 43]. Most of these techniques rely on algorithms that can also be extended to binaural schemes, i. e. using information from both devices [44, 45, 6], in the case of bilaterally-fitted patients.

Unlike other problems that hearing devices attempt to solve, which are mainly caused by the actual patient's impairment, the presence of feedback is caused by the hearing device itself. More specifically, feedback originates when there exists a coupling between the hearing device's components responsible for signal acquisition and signal reproduction. In the standard case, e. g. in a conventional HA, the coupling occurs between the microphone and the loudspeaker of the HA [46]. To a lesser extent, however, feedback might also occur with other

types of hearing devices, such as with BCDs [47] or with the Codacs™ DACI [48].

In order to understand the physical mechanisms causing feedback in these different types of hearing devices, it is important to specify the different types of undesired couplings (i. e. feedback) that can arise in a hearing device. In this thesis, the following four-way classification of feedback will be considered [29]: acoustic, structure-borne mechanical, electrical, and electromagnetic.

- Acoustic feedback refers to a coupling of acoustical nature, i. e. when the transmission between the actuator and the microphone occurs through acoustical waves travelling through air.
- Structure-borne mechanical feedback refers to a coupling of mechanical nature, i. e. when mechanical vibrations are transmitted from the hearing device's actuator to the microphone, through the hearing device case [49, 50, 20, 29]. In conventional HAs, the problem is usually related to a malfunction of the device [49]. The same problem is, however, more relevant in other types of hearing devices, like BCDs [50, 20, 29], where the vibrations of the electromechanical actuators are much more sustained than those of a moving coil loudspeaker.
- Electrical feedback refers to a coupling of electrical nature, i. e. when amplified electrical signals leak to an earlier stage of the amplifier [49, 29].
- Magnetic feedback refers to a coupling of magnetic nature, i. e. when stray electromagnetic radiation caused by magnetic leakage is transmitted from the actuator to earlier parts of the chain in the sound processor [49, 29].

Although different in nature, all these four types of feedback components lead to undesired acoustic artifacts [49, 51, 29], when the gain provided by the hearing device reaches, or in the worst case exceeds, the maximum stable gain (MSG) value. Due to the existence of these different types of feedback components, the common use found in the literature of the adjective *acoustic* to characterize the noun *feedback* may be slightly misleading in some of the situations discussed in this thesis.

The main acoustic effect caused by an unstable behavior within a feedback system is the so-called howling, i. e. the presence of one or multiple high-intensity resonances. Furthermore, a spectrum of subtly diverse artifacts can be identified by many different descriptive words, e. g. "ringing", "screeching", or "buzzing" [49].

Figure 1.3a provides a schematic illustration of the feedback problem on a generic hearing device. The hearing device should provide a properly

compensated version of the desired source signal $v(t)$ to the user. Ideally, $v(t)$ is picked up by the microphone, resulting in the microphone signal $y(t)$. The microphone signal $y(t)$ is subsequently processed through the sound processing stage $G(q,t)$ of the hearing device, also known as the forward path, in order to compensate for the user's hearing impairment. The compensated signal $u(t)$ is then provided to the user through the hearing device output actuator (e. g. an electroacoustic actuator in a conventional HA or an electromechanical actuator in a BCD or a DACI [47]).[4] If a coupling $F(q,t)$ exists between the output actuator and the microphone, i. e. a feedback path, the system is operating in a closed loop and it might exhibit an unstable behavior. Mathematically, the stability of the block diagram in fig. 1.3a can be studied through the Nyquist stability criterion [46]. The Nyquist stability criterion provides a necessary and sufficient condition for instability, through the study of the magnitude and the phase of the open-loop response $G(\omega,t)F(\omega,t)$, i. e.

$$\begin{cases} |G(\omega,t)F(\omega,t)| \geq 1 & (1.1) \\ \angle G(\omega,t)F(\omega,t) = n2\pi, \quad n \in \mathbb{Z} & (1.2) \end{cases}$$

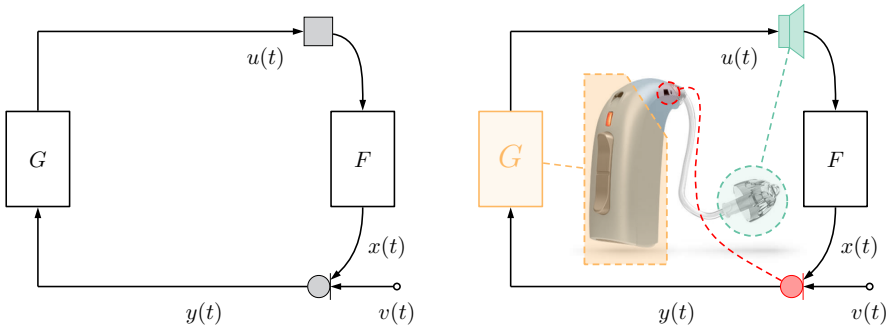
where $G(\omega,t)$ and $F(\omega,t)$ are the short-time Fourier transforms (STFTs) of the forward and the feedback path, respectively, and ω is the radial frequency. At the frequencies where the criterion is fulfilled, the feedback signal $x(t)$ is in phase with and has a greater amplitude than the microphone signal $y(t)$. Therefore, this gives rise to a constructive interference that leads to the well-known feedback artifacts.

Section 1.3.1 and section 1.3.2 provide a description of the feedback problem in HAs and in implantable hearing devices, respectively.

1.3.1 Feedback in hearing aids

In HAs, feedback originates when there exists an undesired coupling between the loudspeaker and the microphone. This coupling creates a closed loop prone to unstable behavior when the forward path gain, i. e. the amplification provided by the HA, exceeds the MSG value. Even though different feedback components might be involved, acoustic feedback is the most common component usually considered in HA literature [51]. This is due to the fact that other feedback components amount to lower contributions compared to the acoustic one and are, usually, related to a malfunction of the HA [49]. Therefore, for simplicity, acoustic feedback alone will be considered when talking about HAs. For a schematic representation see fig. 1.3b.

[4] The quantities t and q^{-1} are the discrete time index and the delay operator, i. e. $q^{-k}u(t) = u(t-k)$, respectively.



(a) Feedback in a hearing device.

(b) Acoustic feedback in a HA.

Figure 1.3: (a): Simplified scheme depicting a general case of feedback. The output actuator could be a loudspeaker or a mechanical actuator. (b): Simplified scheme depicting a case of acoustic feedback in a HA overlayed with a real HA, indicating the forward path, the loudspeaker and the microphone.

As previously mentioned, the most common artifact caused by a system subject to acoustic feedback is howling. This and other related artifacts contribute to lowering the potential satisfaction of a patient fitted with a HA [52]. Therefore, the inclusion of an acoustic feedback control has been positively rated in multiple recent surveys of HA users [53, 54].

In a conventional HA scenario, the acoustic feedback path is dependent on many parameters, some of which are static, such as the type of HA, the vent size, or the actual shape of the outer ear of the patient, and some of which are time-varying or highly time-varying such as the acoustics outside the ear or presence of obstacles around the HA (e. g. hands, a hat, or a smartphone[5]) [49, 55, 56, 57, 43]. The vent design is a typical trade-off case between two undesirable outcomes [51]: on the one hand, a small or nonexistent vent leads to the occlusion effect, i. e. the low-passed and the uncomfortable increased-loudness perception of self-generated sounds, due to the increased perception of bone-transmitted sound [58, 59, 51, 60]; on the other hand, a large venting leads to acoustic feedback [61]. Between the two problems, an algorithmic solution of the acoustic feedback problem seems to be the most feasible.

[5]Although the proximity between the fitted HA and a phone, or a phone handset, is regarded as a common problem, it might become less common a problem in the future as the ability of hearing devices to wirelessly communicate with smartphones improves. This will, in fact, allow the whole audio signals to be streamed to the hearing device without the requirement to bring the smartphone close to the ear.

1.3.2 Feedback in implantable hearing devices

Feedback has been traditionally part of the issues addressed by conventional HAs. However, some implantable hearing devices can also be affected by the problem.

This is the case, e. g., in different types of BCDs such as percutaneous bone-conduction devices (PBCDs) and transcutaneous bone-conduction devices (TBCDs). The technology in PBCDs, such as the BAHA, has a longer history and, hence, the problem of feedback is more extensively documented in the literature [50, 20, 29, 62, 22, 63, 64]. More recently, the feedback problem has also been analyzed in a new TBCD, the BCI, where it does exist at a much lower extent, with 20 dB to 30 dB improvements in the gain margin compared to a PBCD [29].

Although with a much lower incidence, and possibly related to implantation problems and/or improper fitting, there have been complaints about feedback problems even among some of the patients that received a MEI, e. g. the Esteem [65], the Carina [66, 67, 10], and the MAXUM [15], or a DACI, e. g. the Codacs [48]. This suggests that these other types of hearing implants might also be susceptible to feedback.

Unlike in conventional HAs, where acoustic feedback is the most common feedback component, other non-acoustic feedback components become more relevant in implantable hearing devices, such as structure-borne mechanical, electrical, and magnetic feedback components [49, 51, 29], cf. fig. 1.4.[6] Given the different nature of the feedback problem in implantable hearing devices, it is relevant to be able to distinguish the different feedback components. Having more information on the type of feedback involved in the problem may, in fact, simplify the process of reducing the feedback, either by introducing some detail modifications in the design of the device or by better designing an effective AFC algorithm.

For instance, in a recent BAHA model, the Cochlear™ BAHA® BP110 power, a new mechanical design that further physically isolates the electromechanical actuator from the microphones was introduced in order to address the feedback problem [22]. Similarly, adaptive algorithms exploiting prior knowledge on the feedback path might also be employed to ensure faster convergence and lower

[6] In [68], the *structure-borne mechanical* and the *acoustic* feedback components were referred to as *tissular* and *airborne* feedback components, respectively. It was later decided that the words *structure-borne mechanical* and *acoustic* better characterize the feedback components, as they have already been used in literature [29], and will, therefore, be used throughout the thesis. However, the term *acoustic* is preferred over *acoustical*, suggested in [29], in an attempt to follow the guidelines in [69].

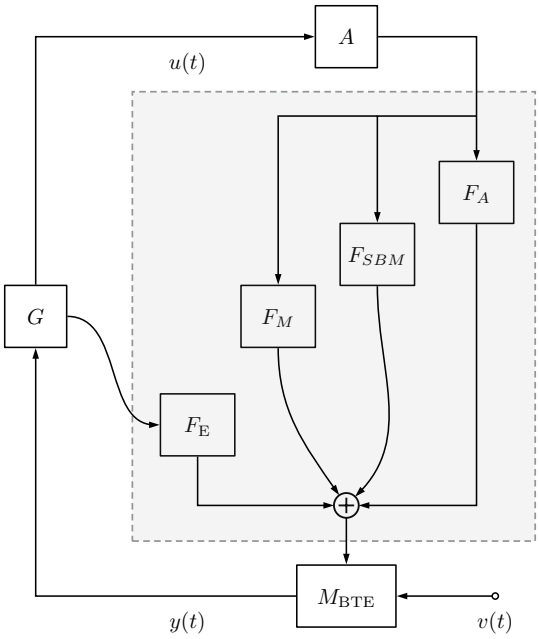


Figure 1.4: Simplified block diagram of an implantable hearing device operating in closed loop. The device is schematically represented by the three blocks $M_{\mathrm{BTE}}(\omega)$, $G(\omega)$, and $A(\omega)$, i. e. the three frequency responses of the BTE microphone, the functions of the sound processor and the RF link, and the actuator, respectively. The four feedback frequency responses, $F_{\mathrm{SBM}}(\omega)$, $F_{\mathrm{A}}(\omega)$, $F_{\mathrm{E}}(\omega)$, and $F_{\mathrm{M}}(\omega)$, indicate, respectively, the structure-borne mechanical, the acoustic, the electrical, and the magnetic feedback. The gray shading summarizes the total feedback frequency response $F(\omega)$. The block diagram has been adapted from fig. 5 in [29].

misadjustment [70, 71, 72].

1.4 Feedback control in hearing devices

The feedback problem in acoustical scenarios has been widely studied in the past five decades. Initially, it has been studied from a purely electronic and acoustical perspective with the early work of Schroeder at Bell Labs, e. g. [73, 74], and, subsequently, from a purely signal processing perspective. Much of the older literature mainly focused on acoustic feedback control for PA systems; however,

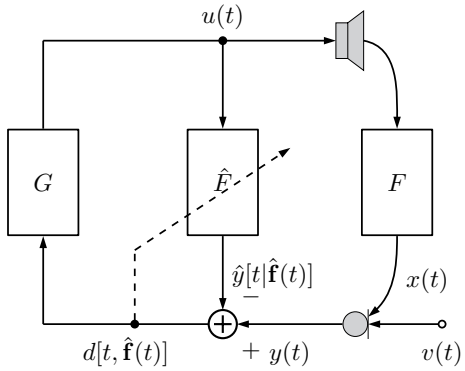


Figure 1.5: General AFC scenario in a HA.

in the past three decades, and especially after the introduction of the first fully-digital HA in 1995, the amount of research focusing on feedback control for HAs has been growing considerably. The main reason of this growth is due to the fact that before the introduction of digital HAs, feedback could mainly be controlled mechanically, as analog and programmable analog implementations of adaptive filters were less flexible to different acoustic scenarios [5].

After an extensive literature review, a four-way classification of these different methods, with a focus on acoustic feedback control for PA systems but that can also be applied to hearing devices, has been proposed in [46]. The four classes include phase-modulation methods, gain reduction methods, spatial filtering methods, and room modeling methods. Many of the solutions developed to tackle the acoustic feedback problem in a HA context have been drawn from the different classes. Most notably, gain reductions methods, e. g. notch-filter-based methods [75, 61, 76, 77, 78], and room modeling methods [79, 80, 81, 82, 83, 84, 85, 86, 87, 88] have been extensively used in HAs. In this thesis, the focus will be solely on one room modeling method, based on AFC algorithms, cf. fig. 1.5.

A review of the state-of-the-art AFC algorithms is given in section 1.4.1. The issue of assessing the performance of AFC algorithms is discussed in section 1.4.2.

1.4.1 Adaptive feedback cancellation algorithms

An AFC algorithm employs the tools provided by adaptive filter theory to remove the coupling between the loudspeaker and the microphone signal.[7] In detail, this is done by including the adaptive filter $\hat{F}(q,t)$, which attempts to estimate the unwanted (and unknown) feedback path $F(q,t)$. Once the adaptive filter $\hat{F}(q,t)$ has been estimated, it can be used to filter the loudspeaker signal $u(t)$ in order to obtain $\hat{y}[t|\hat{\mathbf{f}}(t)]$, i. e. an estimate of the feedback signal $x(t)$. The microphone signal $y(t)$ is modeled as the summation of the unwanted feedback signal $x(t)$ and the desired source signal $v(t)$, i. e. $y(t) = F(q,t)u(t) + v(t)$ assuming a noiseless, linear model for both the microphone and the loudspeaker. Therefore, as the estimate $\hat{F}(q,t)$ gets closer to $F(q,t)$, and hence $\hat{y}[t|\hat{\mathbf{f}}(t)]$ gets closer to $x(t)$, the unwanted feedback signal component $x(t)$ of the microphone signal $y(t)$ can be effectively removed by simply subtracting $\hat{y}[t|\hat{\mathbf{f}}(t)]$ from $y(t)$.

As the adjective *adaptive* suggests, the estimation of $\hat{F}(q,t)$ is carried out adaptively since the unknown feedback path might be highly time-varying, cf. section 1.3.1. This is also why a clear dependence on the discrete time instant t is indicated in $F(q,t)$. The classical and most popular adaptive filtering techniques to provide an adaptive least squares (LS) estimate of the feedback path are the least mean squares (LMS) algorithm [61, 89, 90], its normalized version, the normalized least mean squares (NLMS) algorithm [91, 92, 89, 93, 70, 94, 46, 95], which reduces the sensitivity to the scaling of the loudspeaker signal $u(t)$ [90], or other modifications such as the proportionate NLMS (PNLMS) algorithm [96]. Another very popular adaptive filtering technique that can be used for this purpose is the recursive least squares (RLS) algorithm [83, 97, 98, 99], including an exponential forgetting factor to cope with time variations, especially considering its improved convergence properties over the LMS algorithm [90]. Even though the RLS algorithm has both a much higher computational complexity and poorer numerical properties compared to the LMS algorithm, several fast and numerically stable versions have been derived, e. g. [100, 101]. Nevertheless, these fast RLS algorithms rely on assumptions that are not easily met in an AFC context [46] and this is, possibly, why fewer RLS-based approaches have been used in HA AFC.

Several solutions based on adaptive filters have been successfully applied to acoustic echo cancellation (AEC), a conceptually very similar problem to AFC. However, applying a standard adaptive filter for feedback control is usually not enough for two main reasons:

[7] Although most of the concepts presented in the current section are valid for most kinds of hearing devices, the application case to HAs is considered. That is why the output actuator is referred to as loudspeaker and a loudspeaker symbol is depicted in fig. 1.5.

1. The system continuously operates in a double-talk situation, i. e. the loudspeaker signal $u(t)$ and the source signal $v(t)$ are simultaneously active leading to a slower convergence of the solution;
2. The high correlation between the loudspeaker signal $u(t)$ and the source signal $v(t)$ invalidates one of the fundamental assumptions necessary for the derivation of the adaptive filter.

The second problem, leading to a biased estimation of the feedback path [46], has been addressed in several different ways and a reduced correlation has been obtained, e. g., by introducing an external probe noise [102, 103], and by modifying the forward path $G(q,t)$ by means of nonlinear processing [104, 105], time-varying processing [104, 106] and added delays [92]. The first problem has been essentially disregarded, leading to AFC algorithms with poor convergence behavior.

A promising approach to solve both problems came with the introduction of the so-called prediction-error method (PEM), based on concepts from closed-loop system identification [107, 108], leading to the prediction-error method (PEM)-based AFC algorithms in the early 2000s. The PEM was first included to reduce the bias problem in AFC [97, 82, 83, 93], effectively addressing also the double-talk problem, and subsequently applied to AEC to discard the use of a double-talk detector [99].

Many different PEM-based AFC algorithms have been proposed ever since, both implemented in the time domain [97, 82, 83, 93, 109, 86, 110] and in the frequency domain [84, 110]. Frequency-domain algorithms have the advantage of a lower computational complexity over the time-domain counterparts [111, 112, 113] and are, therefore, appealing in low-power hearing-device applications. Another advantage of frequency-domain adaptive filters (FDAFs) is given by the possibility of using a frequency-dependent stepsize for the adaptive procedure, leading to an improved convergence [114, 113]. However, implementations in the frequency domain introduce longer latencies compared to implementations in the time domain, which might be problematic when the algorithms are expected to run in real time. For this reason, low-latency versions of the FDAF have been developed, such as the partitioned-block frequency-domain adaptive filter (PBFDAF) [115, 98, 116, 84].

1.4.2 Performance assessment of AFC algorithms

In a simulation environment, the performance assessment of AFC algorithms is usually done using two objective measures. The first objective measure is the misadjustment, or mean-square deviation, defining the normalized distance

between the true feedback path and the estimated feedback path [90].[8] The second objective measure is the MSG, i. e. the gain that can be provided to the system while keeping it stable [76, 75, 46]. Alternatively, in place of the MSG, the increase of the MSG upon the adaptive filter insertion, i. e. the added stable gain (ASG) can be used [76, 84]. In order to calculate both the misadjustment and the MSG, or ASG, the knowledge of the unknown feedback path is required, making, hence, these objective measures mainly useful in a simulation scenario. Nevertheless, estimates of these quantities based on the specific AFC algorithm can be obtained [117, 92, 82].

A few recent studies, e. g. [118, 119, 120], have tackled the problem of objectively evaluating AFC algorithms in a more general way. The main scope of these studies has been that of providing a black-box method to objectively evaluate the AFC algorithms of different commercial HAs, as well as novel AFC algorithms. Different objective measures probing several performance aspects related to the tested AFC algorithms, such as instability (e. g. the already mentioned MSG and ASG), oscillation, and distortions, were assessed and found to be usable to provide a meaningful evaluation of the algorithms under test.

Even though these objective measures are helpful to obtain a performance estimate for different aspects related to the tested AFC algorithms, they cannot be directly used to infer the performance with respect to an important perceptual aspect such as the sound quality. For this reason, in recent years, objective measures to estimate more perceptually meaningful audio attributes have been introduced for AFC algorithms evaluation [121, 122]. It must be mentioned that, compared to other tasks in the field of audio signal processing, such as noise reduction or audio coding, the number of studies dealing with the performance assessment of AFC algorithms from a perceptual point of view is very limited, e. g. [76, 89, 121, 122, 123, 124, 125, 126, 118, 119, 120]. Furthermore, within this limited set of studies, only a handful lead to the development of sound-quality objective measures validated in a feedback scenario and are, hence, usable to assess the performance of AFC algorithms. These sound-quality objective measures are the frequency-weighted log-spectral signal distortion (SD), validated through a real-time evaluation [121, 118, 119, 120], the HA speech quality index (HASQI), validated through a subjective evaluation [122], and a combination of different existing sound-quality measures combined by means of a neural network, also validated through a subjective evaluation [123]. Of the three mentioned objective measures, the SD has been used more extensively due to its simplicity and the earlier introduction [121, 46, 127]. Other common sound-quality objective measures have also been employed to evaluate the sound quality resulting from different feedback control algorithms

[8] In this thesis, the full time evolution of the misadjustment is considered. Alternatively, one could focus on the convergence rate, the steady-state error or the tracking error [117].

(i. e. not only AFC algorithms), e. g. the perceptual evaluation of speech quality (PESQ) [128, 129, 102, 130, 131] and the perceptual evaluation of audio quality (PEAQ) measures [123, 102, 130]. However, this was done even though neither the PESQ nor the PEAQ measure were specifically designed with the scope of evaluating changes in the sound quality caused by acoustic feedback control algorithms.

Overall, the choice of using an objective measure over a subjective one is done for a matter of simplicity and costs, due to the difficulties in performing subjective evaluations [132]. Nevertheless, an objective measure should be used as a predictor only after validating it on the considered task.

1.5 Overview of the thesis

The three parts of this thesis present as many different tasks related to the development of a feedback control strategy for a novel hearing device.

1.5.1 General overview

The **first part** of this thesis (chapters 2 and 3) focuses on the feedback characterization of two novel hearing devices, the Cochlear™ Codacs™ DACI and an early prototype of a BCIC. The measurements have been performed on fresh frozen cadaver heads and are used to investigate different important aspects of the feedback these two implants may experience, such as specimen-dependent behavior, nonlinearities, gain margins, and effects of structure-born mechanical versus acoustic feedback.

The **second part** of this thesis (chapters 4 and 5) describes the two contributions on the topic of AFC, i. e. two novel AFC algorithms. The first is an all-frequency-domain method of the PEM-based frequency-domain adaptive filter (PEM-FDAF), i. e. the frequency-domain prediction-error-method-based adaptive filter (FD-PEMAF), relying solely on frequency-domain operations. The second is a PEM-based AFC algorithm replacing the standard adaptive filter with a simplified Kalman filter, i. e. the PEM-based frequency-domain Kalman filter (PEM-FDKF) algorithm.

The **third part** of this thesis (chapter 6) presents a study involving subjective listening tests investigating the sound-quality degradation caused by different AFC algorithms, including the PEM-FDKF. Additionally, the subjective results are compared to the performance obtained using a batch of different objective measures.

1.5.2 Thesis outline

Chapter 2 presents the feedback path characterization measurements performed on the Cochlear Codacs DACI. The data analysis on the collected measurements reveals multiple findings. Namely, the specimen-dependency of the feedback path, the strong effects of an incorrect implantation, the nonlinear behavior at high stimulus levels and, finally, the dual nature of the feedback path composed by both a structure-born mechanical feedback component and an acoustic feedback component.

Chapter 2 has been published as:

- **G. Bernardi**, T. van Waterschoot, M. Moonen, J. Wouters, M. Hillbratt, and N. Verhaert, "Measurement and analysis of feedback and nonlinearities for the Codacs direct acoustic cochlear implant," *IEEE Access*, vol. 5, pp. 8702–8713, May 2017.

Chapter 3 presents the feedback path characterization measurements performed on an early prototype of a BCIC. *Chapter 3 cannot be shared publicly due to confidentiality obligations.*

Chapter 4 presents the derivation of the frequency-domain prediction-error-method-based adaptive filter (FD-PEMAF). This AFC algorithm is similar to the PEM-FDAF, but includes a frequency-domain implementation of the prefiltering stage, thus performing all the filtering operations and the filter adaptation in the frequency domain. Such an approach allows the lowering of the computational complexity while maintaining comparable performance.

Chapter 4 has been published as:

- **G. Bernardi**, T. van Waterschoot, J. Wouters, and M. Moonen, "An all-frequency-domain adaptive filter with PEM-based decorrelation for acoustic feedback control," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA '15)*, (New Paltz, NY), pp. 1–5, Oct. 2015.

Chapter 5 presents the derivation of the PEM-FDKF, a frequency-domain AFC algorithm replacing the more common FDAF with a diagonalized frequency-domain Kalman filter (FDKF). A PEM stage is also included to limit the estimation bias. Both the fullband version, i. e. the PEM-FDKF algorithm, and the subband version, i. e. the PEM-based partitioned-block frequency-domain Kalman filter (PEM-PBFDKF) algorithm, are presented. Additionally, the identifiability conditions for the optimization problem leading

to the proposed PEM-FDKF algorithm are derived. The PEM-PBFDKF algorithm combines good decorrelation properties, by means of the PEM-based signal prewhitening, fast convergence, good tracking behavior, and low computational complexity by means of the FDKF, and low delay by means of a partitioned block (PB) implementation.

Chapter 5 has been published as:

- **G. Bernardi**, T. van Waterschoot, J. Wouters, and M. Moonen, "Adaptive feedback cancellation using a partitioned-block frequency-domain Kalman filter approach with PEM-based signal prewhitening," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, pp. 1480–1494, Sep. 2017.

Chapter 6 presents the subjective evaluation of the PEM-FDKF, by means of a listening test. The subjective evaluation suggests that the PEM-FDKF yields a higher sound quality compared to other AFC algorithms. Additionally, the subjective results are compared with the results of existing objective scores, showing that a good subjective-to-objective correlation can be obtained even when using some objective scores not specifically designed to evaluate the sound-quality degradation by acoustic feedback artifacts.

Chapter 6 has been submitted for publication as:

- **G. Bernardi**, T. van Waterschoot, J. Wouters, and M. Moonen, "Subjective and Objective Sound-Quality Evaluation of Adaptive Feedback Cancellation Algorithms," Submitted for publication to *IEEE/ACM Trans. Audio Speech Lang. Process.*, Nov. 2017.

Chapter 7 concludes the thesis by reiterating the different objectives and results presented in the different chapters and exploring some directions for future research.

Part I

Feedback Characterization

Chapter 2

The Codacs direct acoustic cochlear implant

Measurement and Analysis of Feedback and Nonlinearities for the Codacs Direct Acoustic Cochlear Implant

Giuliano Bernardi, Toon van Waterschoot, Marc Moonen, Jan Wouters, Martin Hillbratt, and Nicolas Verhaert

Published in *IEEE Access*, Vol. 5, pp. 8702–8713, May 2017, DOI: 10.1109/ACCESS.2017.2703635.

© 2017 IEEE. Reprinted, with permission, from:
G. Bernardi, T. van Waterschoot, M. Moonen, J. Wouters, M. Hillbratt, and N. Verhaert, "Measurement and analysis of feedback and nonlinearities for the Codacs direct acoustic cochlear implant," *IEEE Access*, vol. 5, pp. 8702–8713, May 2017.
Changes include layout, representation, and minor editing aspects.

The candidate's contributions as first author include: literature study, co-design of the measurement protocol, co-participation in the measurement sessions, software implementation of the measurement technique and of the analysis tools for the measured data, co-analysis of the measured data, co-formulation of the conclusions, text redaction and editing.

Abstract

Acoustic feedback is a very common problem in hearing instruments. Not only does it occur in common behind-the-ear (BTE) or in-the-ear hearing aids, it also affects bone conduction implants, middle ear implants, and more recent devices, such as the direct acoustic cochlear implant (DACI). In this paper, we present the data and analysis relating to the feedback path characterization of the Cochlear™ Codacs™ DACI, performed on fresh frozen cadaver heads in four different measurement sessions (MSs). The general objectives were the following: 1) To measure and analyze the feedback path of the system and check for possible specimen-dependent variabilities; 2) To assess whether this feedback path is affected by an incorrect implantation; 3) To check for nonlinear behavior and 4) To determine differences between structure-borne mechanical and acoustic feedback.

The data analysis reveals that the feedback seems to be dependent on the specific head morphology of the implanted specimen, and that an incorrect implantation might strongly affect the feedback path; additionally, the analysis reveals that some nonlinear behavior at high stimulus levels can be expected and, finally, that the feedback path is characterized by a structure-borne mechanical feedback component with a rather different frequency content compared to the acoustic feedback component.

2.1 Introduction

Hearing loss occurs when some form of impairment affects the auditory system. Given the complex structure of the auditory system, a broad spectrum of problems can affect normal hearing. For this reason, in the last five decades, several types of hearing instruments have been designed to cope with these different hearing loss conditions. Various solutions have been specifically designed for both conductive hearing loss and sensorineural hearing loss, while mixed hearing loss, in particular when severe to profound, could not be addressed with a dedicated solution until the last decade. One of the implants able to cope with such a problem is the so-called direct acoustic cochlear implant (DACI) [33], see fig. 2.1.

The Cochlear™ Codacs™ DACI consists of an implantable electro-mechanical actuator, placed within the mastoid cavity and firmly fixed through a bone plate, see fig. 2.1b and fig. 2.1c, directly stimulating the cochlea via a piston prosthesis coupled to an artificial actuator [3]. The implanted actuator is controlled and driven from the outside through an RF link, and can be interfaced with the

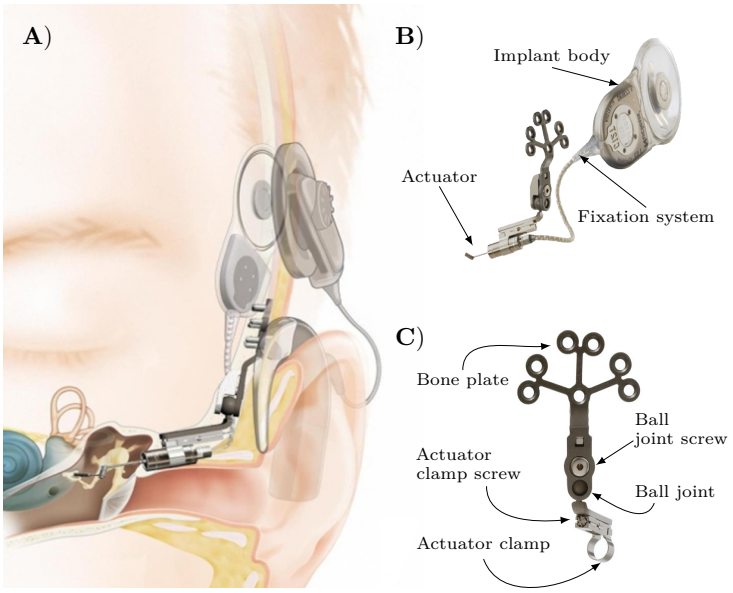


Figure 2.1: Details of the Cochlear™ Codacs™ DACI. A) Scheme showing both the implanted and the external parts of the device. B) Implantable parts. C) Fixation system. Courtesy of Cochlear Ltd.

standard Cochlear behind-the-ear (BTE) sound processor (originally developed for cochlear implants). The Codacs DACI has been proven to be effective in the treatment of severe-to-profound mixed hearing loss [133]; however, the follow-up monitoring of some patients that received the implant suggested that acoustic feedback can occur [48].

In order to tackle the feedback problem, some kind of feedback control is needed; a widely used strategy to control acoustic feedback involves the use of an adaptive feedback cancellation (AFC) algorithm, where a filter adaptively estimates the unknown feedback path and hence the unknown feedback signal, which is then subtracted from the microphone signal [46, 76]. The design of an AFC algorithm can be simplified if some prior knowledge of the feedback path is available; for instance, from single or multiple characterizations of the feedback path, one can evaluate the presence of nonlinearities, the time- and frequency-domain structure of the feedback path, as well as tune some critical parameters of the AFC algorithm, such as the filter length and the overall scaling factor of the filter coefficients [61, 70]. Additionally, the variability introduced by changes in the hearing aid (HA) working conditions, e. g. due to a hand or a

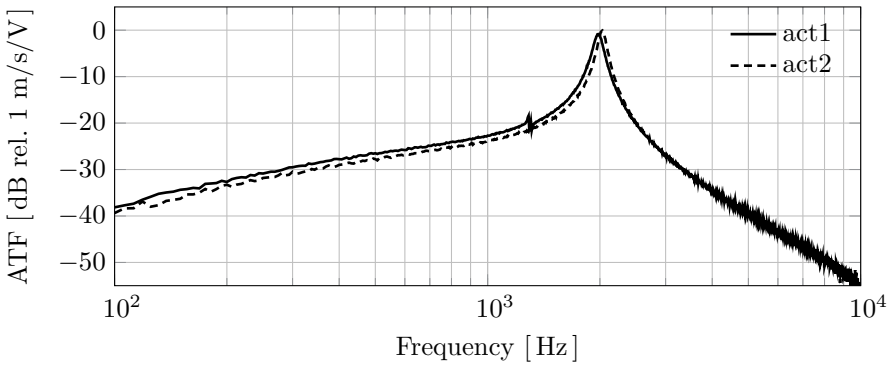


Figure 2.2: Vibration velocity to voltage sensitivity as a function of frequency for the two Codacs actuators (`act1` and `act2`) used throughout the four MSs.

different reflector nearby the HA microphone, can be evaluated [79, 119].

A characterization of the feedback path, either on test subjects or mannequin heads, using random noise signals is often performed in the literature introducing AFC and other feedback-control algorithms for HAs applications [55, 75, 61, 91, 79]. The characterization measurements presented in this paper, instead, were carried out on fresh frozen cadaver heads; this was done because the Codacs DACI is an implantable device, and hence cannot be mounted on a mannequin head, and because any risk for the implanted patients deriving from excessive stimulation needed to be avoided. Additionally, the characterization was performed using the exponential sine sweep (ESS) technique, allowing to monitor the nonlinear content of the system under test [134].

The measurements and analysis of the feedback path of the Codacs DACI presented in this paper had multiple general objectives. First, the feedback path of this novel hearing instrument was measured, providing information in terms of filter length, levels and spectral shape given that this knowledge can simplify the design of a feedback canceller. The tests were performed on four cadaver heads to verify possible feedback variabilities. Second, the measurements were performed in a case of an incorrect actuator positioning (i. e. the actuator touched the bone of the posterior wall of the external ear canal) to assess whether this could lead to an increase of the feedback energy in possibly critical regions. Third, the presence of nonlinearities was investigated, since these can have a profound impact on the performance of standard feedback cancellers. Finally, it was assessed whether the feedback path impulse response (IR) measured with the microphone positioned on a standard BTE sound processor differs from the IR measured with an external (and mechanically decoupled) microphone.

This is done to verify whether the mechanical coupling between the DACI and the BTE microphone, through the bone tissue and the soft tissue layers, gives rise to a significant non-acoustic, referred to as structure-borne mechanical, feedback component. The term mechanical coupling is used loosely to indicate the mechanical interactions between the DACI and the BTE microphone, caused mainly by the fixation of the DACI to the temporal bone and, to a lesser extent, by the connection between the DACI and the cochlea. Thus, by structure-borne mechanical feedback we mean the acoustic feedback propagating mainly through bone and soft tissue layers and not through air.

2.2 Materials and methods

2.2.1 Codacs actuator transfer function

The Codacs DACI is characterized by a strong resonance in the actuator transfer function (ATF) [135]; the ATF is the transfer function measured on the bench between an applied voltage and the induced stapes prosthesis velocity and it is expressed in m/s/V [4]. Laser-Doppler-vibrometry measurements show that the ATF exhibits a resonance around 2 kHz, that can slightly shift when using different actuators. In implanted patients, a more relevant source of variability is due to implantation and postoperative processes such as healing and scarring. The impact of these phenomena is discussed later in section 2.4.1. The effect of the resonance is usually precompensated by filtering the signal fed to the actuator with the inverse ATF.

Two Codacs actuators, referred to as `act1` and `act2`, were used in the different measurement sessions (MSs) (`act1` in MS1, MS2 and MS3, `act2` in MS4); their frequency responses are shown in fig. 2.2.

2.2.2 Feedback system description

Figure 2.3 depicts the simplified block diagram of an implanted Codacs DACI, operating in a closed-loop scenario. This does not represent the tested system described in this paper, but it is helpful to contextualize how a characterization of the feedback path of the implanted device can be helpful to the design of an AFC algorithm. Each block in the scheme is represented via its frequency response, defined as a function of the angular frequency ω. The rightmost block $F(\omega)$ represents the feedback path frequency response, expressed in m/s/Pa, relating the actuator velocity to the pressure level recorded by the microphone. The Codacs DACI is schematically represented by three main parts: the BTE

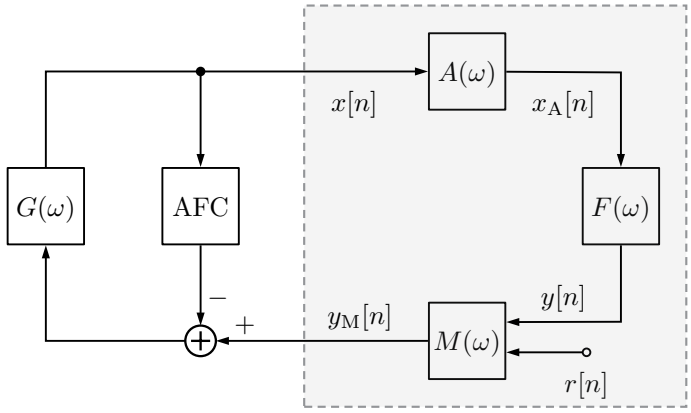


Figure 2.3: Simplified block diagram of an implanted Codacs DACI operating in closed-loop , where $M(\omega)$, $A(\omega)$, $G(\omega)$ and $F(\omega)$ represents the frequency responses of the BTE microphone, the Codacs actuator, the sound processor, and the feedback path, respectively, AFC represents the AFC algorithm, and $r[n]$ represents the source signal (such as speech or background noise). The shaded portion of the diagram refers to the open-loop system, schematizing the measurement scenario of the different MSs.

microphone characterized by $M(\omega)$, the Codacs actuator characterized by $A(\omega)$, and the sound processor characterized by $G(\omega)$.

At high stimulation levels, the actuator is assumed to be a weakly nonlinear system, as often done for loudspeakers, and can be modeled by means of memoryless Volterra series [134, 136]. In a hearing aid scenario $G(\omega)$ is often including some kind of nonlinear processing (e. g. compression), too. The frequency responses $M(\omega)$, and $F(\omega)$ are assumed to be linear and time-invariant, meaning that they can be all replaced by means of the correspondent IRs, i. e. $m[n]$ and $f[n]$. In such a case, the microphone signal $y_{\mathrm{M}}[n]$ is obtained from the addition of two signals: $r[n]$ is the source signal given by the desired component to process and amplify, such as speech or music, and possibly some background noise; $y[n]$ is the feedback signal, obtained by filtering the stimulus $x_{\mathrm{A}}[n]$ generated by the Codacs actuator through the feedback path IR $f[n]$, i. e. $y[n] = f[n] * x_{\mathrm{A}}[n]$ where $*$ denotes convolution, contributing to the closure of the feedback loop. Memoryless Volterra series provide a compact way to describe the relation between $x[n]$ and $x_{\mathrm{A}}[n]$ by means of the Volterra kernels, i. e. $a_k[n]$ with $k = 1, \ldots, K$; however, discussing such a relation is outside the scope of this paper. The cascaded feedback path IR can be included in the actuator's model by replacing the Volterra kernels $a_k[n]$ with the modified

Volterra kernels of the system including both actuator and feedback path, i. e. $h_k[n]$ with $k = 1, \ldots, K$, whose estimates will be used later to define the nonlinearity measure. Specifically, the relation $h_1[n] = a_1[n] * f[n]$ holds [137]; under the assumption that the linear Volterra kernel $a_1[n]$ can be approximated by the ATF, a precompensation can be introduced by filtering the input signal $x[n]$ fed to the actuator with an estimate of the inverse $a_1[n]$, i. e. $\hat{a}_1^{-1}[n]$, to obtain the proportionality relation $h_1[n] \propto f[n]$ and hence simplify the identification procedure of $f[n]$.

If the nonlinearity assumption of the actuator is relaxed (this is reasonable at low stimulation levels, as will be shown in section 2.3.3), and the same is done with $G(\omega)$, leading to a linear representation of $A(\omega)$ and $G(\omega)$, the stability of the closed-loop system in fig. 2.3, in absence of the AFC block, can be analyzed [75, 76, 46]. Although assuming linearity for $G(\omega)$ is arguably simplistic in a hearing aid scenario, this is common in literature to keep the theoretical analysis treatable [84, 61, 79, 92]. The product between $F(\omega)$ and the forward path response, represented by $A(\omega)G(\omega)M(\omega)$, results in the so-called open-loop response $F(\omega)A(\omega)G(\omega)M(\omega)$. The analysis of the open-loop response is important since it allows to determine the stability of the closed-loop response from the source signal $r[n]$ to the actuator output signal $x_\mathrm{A}[n]$, by means of their frequency responses $R(\omega)$ and $x_\mathrm{A}(\omega)$, respectively:

$$\frac{x_\mathrm{A}(\omega)}{R(\omega)} = \frac{A(\omega)G(\omega)M(\omega)}{1 - A(\omega)G(\omega)M(\omega)F(\omega)}. \tag{2.1}$$

Therefore, the estimation of $F(\omega)$ discussed in this paper represents an important first step to calculate the open-loop response of the system under test that will be used in the stability analysis and control.

2.2.3 Feedback path estimation

In measurements performed for this study, we used the ESS technique [134], widely used in room impulse response (RIR) measurements [138, 137]. The ESS technique is used to estimate the feedback path IR $f[n]$ and to investigate the nonlinear behavior of the actuator.

A mathematical description of the ESS stimulus is given by

$$x_\mathrm{ESS}[n] = \sin\left\{\frac{\omega_1 T}{\ln(\omega_2/\omega_1)}\left[\left(\frac{\omega_2}{\omega_1}\right)^{\frac{n}{T}} - 1\right]\right\}, \tag{2.2}$$

where ω_1 and ω_2 are the starting and ending angular frequencies of the sweep, and T is the duration of the sweep in s. Hence, the ESS stimulus has a frequency

content that exponentially varies with time with a roughly pink spectrum (i. e. -3 dB/Octave). Usually, a fade-in and fade-out stage at the extremes of the sweep is also included [138]. In the current study, we used a 2000-samples Hann window.

Referring to fig. 2.3, the input and output of our measurements were $x[n]$ and $y_{\mathrm{M}}[n]$, respectively. The actuator was fed with a precompensated stimulus $x[n]$, obtained by filtering $x_{\mathrm{ESS}}[n]$ via the inverse of the available ATFs, i. e. $x[n] = x_{\mathrm{ESS}}[n] * \hat{a}_1^{-1}[n]$. This was mainly done in order to prevent overstimulation and extreme distortion that might be induced by the actuator's resonance, but also to simplify the identification procedure, as seen in section 2.2.2. By doing so, a first order approximation of the input fed to the feedback path $F(\omega)$ is $x[n] * a_1[n] \approx x_{\mathrm{ESS}}[n]$.

The deconvolution procedure to estimate $f[n]$ is easily obtained by filtering the measured output signal $y_{\mathrm{M}}[n]$ with the time-reversed version of $x_{\mathrm{ESS}}[n]$ [134]. The ESS deconvolution results in a signal, from now on referred to as deconvolved total response (DTR), characterized by a sequence of time-separated responses, cf. fig. 2.4, each representing a different-order harmonic distortion. If the system is linear and the linear Volterra kernel $a_1[n]$ is properly compensated for, the sequence consists of a single response proportional to an estimate of the feedback path IR $f[n]$, i. e. $\hat{h}_1[n]$, from now on referred to as deconvolved impulse response (DIR). In case of a nonlinear system, other higher-order harmonics, i. e. $\hat{h}_k[n]$ for $k = 2, \ldots, K$, will appear before the DIR [134]. The DTR can be compactly defined as $\hat{h}[n] = \sum_{k=1}^{K} \hat{h}_k[n]$.

In reality, the measured output signal $y_{\mathrm{M}}[n]$ includes the effect of the microphone transfer function, too, when the latter is not flat. More details will be given in section 2.4.

A sampling frequency $f_{\mathrm{s}} = 44.1$ kHz was employed during the measurements and the stimuli were constructed as follows: The length of the ESS was set to be 8 s and the swept was between 50 and 10 000 Hz. Two seconds of silence were placed between each of the five sweeps repeated for each of the four applied ESS stimulus levels, i. e. 0.18, 0.35, 0.53, and 0.71 $\mathrm{V_{RMS}}$. The highest tested level was chosen considering that the actuator's internal magnetic circuit starts saturating above 0.71 $\mathrm{V_{RMS}}$, and corresponds to an equivalent sound pressure level between 120 and 130 dB SPLeq; the other levels were chosen to be linearly spaced between 0 and 0.71 $\mathrm{V_{RMS}}$. In a patient, the stimulus levels are usually lower than 10 $\mathrm{mV_{RMS}}$ [139], but they can easily reach levels in the order of hundreds of $\mathrm{mV_{RMS}}$ (private communication with the manufacturer). The stimulus levels used in this study were chosen to be higher in order to cope with the low signal-to-noise ratio (SNR) of the measurement room.

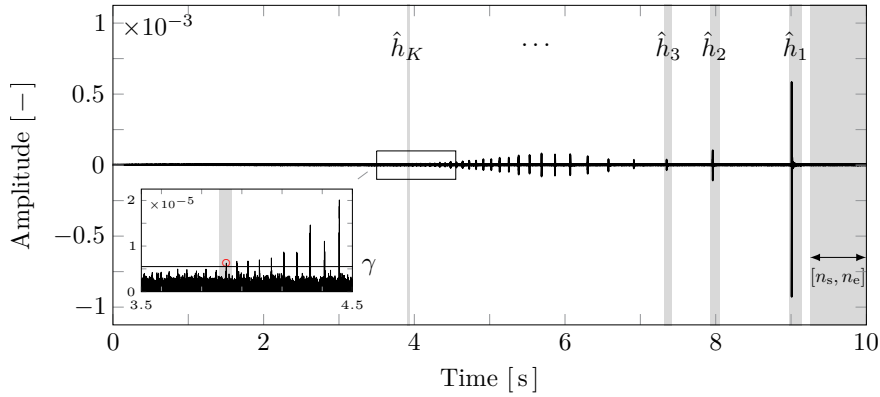


Figure 2.4: Example of a DTR in the time domain from one of the measured BTE microphone signals. The inset shows the amplitude absolute values in the part of the response surrounding the K^{th} harmonic, with ranges 3.5 s to 4.5 s and 0 to 2.5×10^{-5} in the x- and y-axis respectively, and helps to compare the amplitudes of the harmonics and those of the noise floor (i. e. $n \in [n_s, n_e]$).

The choice of retaining only five repetitions of the sweep per measurement was mainly driven by practical constraints, such as the non-ideal acoustic conditions of the room (e. g. the presence of impulsive noise) and the limited time that could be allocated to the use of each cadaver head during the measurements. Due to the difficulties of recording five consecutive sweeps without any impulsive background noise corruption, several recordings were taken and only those with the lowest (or absent) traces of impulsive noise were retained and patched together for the subsequent analysis. Assuming additive and uncorrelated background noise, the coherent averaging of the five recorded sweep responses allowed to reduce the noise floor by approximately $10 \log_{10}(5) \approx 7$ dB.

2.2.4 Distortion measure

A common nonlinear distortion measure employed to characterize (acoustic) nonlinear systems is the so-called total harmonic distortion (THD) [140]. Unfortunately, an extension of the THD to be used with the ESS technique has not been developed yet. Additionally, in our measurements, it was not feasible to carry out an extensive THD characterization measurement, consisting of single tone stimulations repeated at each tested frequency, due to the time constraint imposed by the use of cadaver heads. Therefore, we introduce a nonlinear

distortion measure calculated using only easily distinguishable features from the DTRs.

Figure 2.4 provides an example of a DTR from one of our recordings, highlighting some of the key parameters for the calculation of the proposed nonlinear distortion measure, where the gray shaded areas $\hat{h}_k[n]$ are estimates of the modified Volterra kernels $h_k[n]$. The shaded area $\hat{h}_1[n]$ in the righthand side of the figure approximates the linear modified Volterra kernel, shown to be proportional to $f[n]$, while the other shaded areas $\hat{h}_2[n] \dots \hat{h}_K[n]$ approximate the higher order kernels, related to the 2^{nd} to K^{th} harmonics [134]. K was chosen to be the minimum number of identifiable maxima in the DTR, after a lower threshold value for the maxima was fixed, i. e. no maximum below the threshold was retained. This was done in order to avoid selecting spurious noise peaks with amplitude comparable to the true higher order harmonics. The threshold was estimated from the rightmost portion of $\hat{h}[n]$ in fig. 2.4, i. e. for $n \in [n_\text{s}, n_\text{e}]$, assumed to contain only noise. Such an assumption should be fulfilled, given that n_s was chosen 0.2 s after the beginning of the linear part $\hat{h}_1[n]$ and, as we will see in section 2.3, 0.2 s is approximately one order of magnitude greater than the decay time of the $\hat{h}_1[n]$. Specifically, the threshold was conservatively calculated as $\gamma = \max_{n \in [n_\text{s}, n_\text{e}]} \left\{ |\hat{h}[n]| \right\}$. The inset of fig. 2.4 helps to visually compare the amplitude absolute values of the different harmonics and those of the noise floor samples. The circled relative maximum represents the value of $\max \left\{ |\hat{h}_K[n]| \right\}$, i. e. the lowest supra-threshold value, in the specific situation.

The calculation of THD requires an estimation of the power of each harmonic; however, this is not easily done with our measurements due to the low SNR. Therefore, our distortion measure D is calculated using only the amplitude absolute value maxima of the different harmonics:

$$D = \frac{\sum_{k=2}^{K} \left[\max \left\{ |\hat{h}_k[n]| \right\} \right]^2}{\left[\max \left\{ |\hat{h}_1[n]| \right\} \right]^2}. \tag{2.3}$$

It must be pointed out that this nonlinear distortion measure does not necessarily provide a general indicator of system nonlinearity, comparable to other known measures like THD, but it provides a means to compare the results collected in our measurements from different sessions, using different excitation levels and different implantation schemes.

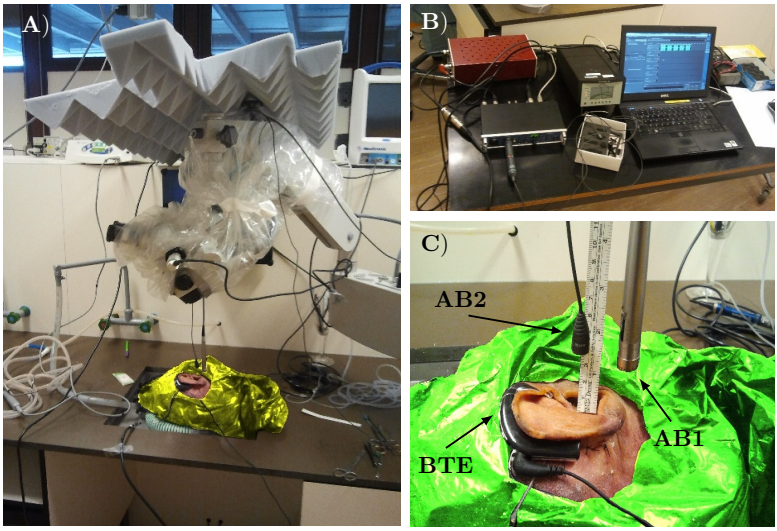


Figure 2.5: MSs setup. A) The cadaver head (right ear) with the implant in place and the three microphones ready to record. B) The instrumentation used for the measurements. C) A close-up of the cadaver head (right ear). In both A) and C), the plastic bag around the head has been digitally colored to differentiate its original black color from the colors of the microphones.

2.2.5 Cadaver heads measurements

All the feedback path measurements on thawed fresh frozen cadaver heads took place in the ENT Department (Saint-Luc University Hospital, Brussels, Belgium) between July 2013 and June 2014 for a total of four MSs and the study was approved by the local ethical committee. Entire heads were evaluated after obtaining authorization to use organs and tissues for research (Science Care, Inc., Phoenix, AZ, USA). The medical history of each head was provided by the supplying company to exclude any otologic pathology. In each of the MSs, a different cadaver head was used and the surgery was performed on the right ear, after any residual earwax had been removed and otomicroscopy verified an intact ear drum and the absence of ear pathology. Surgical preparation was carried out similarly for all four MSs and included a routine Codacs implantation [48]. In detail, it consisted of a canal wall-up mastoidectomy with preservation of the posterior border of the mastoid cavity for placement of the implant's fixation system. A large posterior tympanotomy was performed by opening the facial recess. Care was taken not to touch the ossicular chain avoiding trauma to the ossicles and eardrum. Subsequently, exposure of stapes crurae

and stapes footplate was obtained. After a stapedotomy, allowing to verify that the cochlea was filled with fluid, and fixation of the actuator in the mastoid cavity, a stapes prosthesis was coupled to an actuator. The stapedotomy opening was subsequently sealed with fibrous tissue. No additional middle-ear transfer functions were obtained as with the Codacs surgery the ossicular chain is interrupted. Finally, in order to reproduce real-life implantation, the wound was closed with sutures in a double-layer technique. A small actuator lead was passed through the incision between two sutures to avoid leakage.

The choice of using cadaver heads was made in order to be able to perform the measurements at high stimulation levels, possibly uncomfortable for implanted patients, such that a higher SNR could be achieved. Furthermore, in this way, an incorrect implantation, thought to be a possible cause of the problems highlighted in some of the follow-up cases in living patients, could be tested. The measurements were carried out in a room that was not acoustically treated, due to the practical difficulties of moving the cadaver heads to a separate acoustically shielded measurement room.

The stimuli were digitally created using a laptop, transmitted to a soundcard (RME Fireface UCX) and passed through a laboratory power amplifier (N4L LPA01) which was directly connected to the Codacs actuator. This measurement setup corresponds to direct stimulation in the sense that the sound processor was bypassed. The output signal of the amplifier was also sent back to the soundcard as a control signal to assess that this initial part of the signal path was not introducing any nonlinear distortion. Simultaneously to playing back the stimuli, the acoustic signal at the output of the ear canal were recorded by means of three microphones: a microphone embodied in the BTE case mounted on the cadaver head which, from now on, will be referred to as BTE microphone; a Brüel & Kjær 4190-L-001 with a 2690-A NEXUS microphone conditioner and an AKG CK 97-C, from now on referred to as airborne1 (AB1) and airborne2 (AB2) microphone, respectively. These two external microphones were hung at a distance of roughly 3-to-5 cm from the ear of the cadaver and placed so as to avoid any contact between each other. All the processing stages of the BTE sound processor were bypassed, except for the 12-ms input/output delay. The three microphones were connected to the soundcard, their recordings were digitized, using a 24-bit precision, and transferred back to the laptop in real time. The choice of using two airborne (AB) microphones was done to increase the MS setup redundancy. However, due to the unforseen unreliability of the AB2, exhibiting a level-dependent behavior at low frequency for the tested levels, we focused on the AB1 microphone solely. Real pictures of the setup are shown in fig. 2.5A-C.

Two of the four MSs returned only partial data: In MS3 the AB1 microphone was not available, while in MS4 we did not perform the full set of measurements

carried out in the other three sessions, but only tested the highest stimulus level. Neverthelss, in MS4 we were able to test the differences originating from a correct implantation and a specific type of incorrect implantation of the device. In this paper, incorrect implantation refers to the case in which the tip of the Codacs actuator touched the bone of the posterior wall of the external ear canal. Other types of incorrect implantation could occur but we did not consider them in this work.

2.2.6 Calibration measurement

In addition to the cadaver heads measurements, we performed a microphone calibration measurement, in order to map the recorded levels from the BTE and the AB1 microphones into dB SPL units. We only performed an amplitude calibration, i. e. no phase calibration was performed. The BTE microphone calibration was done using a small anechoic chamber (Brüel & Kjær 4222) together with a Pressure-field 1/2" Brüel & Kjær 4192 reference microphone, and the AB1 microphone calibration was done using a Brüel & Kjær 4230 sound level calibrator.

2.2.7 Data presentation

The time-domain signals shown in this paper are either microphone signals, DTRs or DIRs. The microphone signals are expressed in V, while both the DTRs and the DIRs are dimensionless, see section 2.2.2. In the DIRs plots, we simply set the time origin of our plots in correspondence with the DIRs first peak. This was done due to the small distance between the actuator and the BTE/AB1 microphone (less than 10 cm), which would correspond to a direct-sound delay on the order of tenths of ms. Such a small delay would be negligible compared to the actual processing delay introduced by the sound processor in an implanted patient (at least two orders of magnitude smaller).

The frequency-domain signals shown in this paper are either magnitude responses or power spectral densities (PSDs): the magnitude response of the DIRs were calculated using only the linear part of the DTR, considered to be a deterministic signal; the PSDs were used to qualify the whole calibrated microphone signals, considered to be stochastic signals. Furthermore, when showing the BTE PSDs, the dB SPL levels were forced to $-\infty$ above 7 kHz, to avoid improper scaling due to the strong high-frequency cutoff of the BTE microphone.

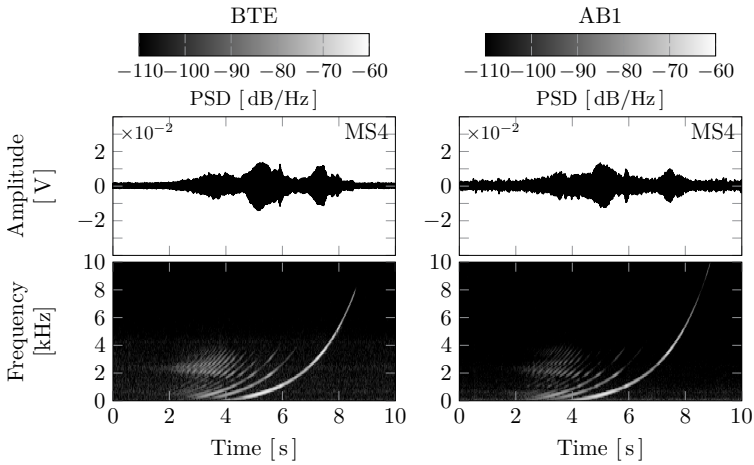


Figure 2.6: Recorded microphone signals and corresponding spectrograms from BTE and AB1 microphones obtained with a $0.71\,V_{RMS}$ ESS stimulus in MS4.

2.3 Results

2.3.1 Cadaver heads measurements

Figure 2.6 shows the raw microphone signals (together with their spectrograms) for both BTE and AB1 microphones when the system was excited with a $0.71\,V_{RMS}$ ESS stimulus in MS4. We notice that the actuator behaves in a nonlinear way, when excited with the highest tested level, as indicated by the visible higher-order harmonics observed in both microphones signals (no copies of the original sweep should be visible if the system were linear). The frequency region at approximately $2600 \pm 400\,Hz$ is excited at times when the stimulus frequency ranges roughly from 80 to 500 Hz. Therefore, the excited harmonics are of rather high order (at least 35 harmonics, on a visual analysis). For the sake of brevity, the microphone signals from the other MSs and the other levels are not shown. However, we would like to point out that the nonlinearity pattern varies but preserves the same core structure in all the different measurements.

Figure 2.7 shows the DIRs for both BTE and AB1 microphones when the system was excited with a $0.71\,V_{RMS}$ ESS stimulus, for all four MSs. The AB1 microphone recordings from MS3 are missing in fig. 2.7 (as well as in figs. 2.9 to 2.11) due to the non-availability of the AB1 microphone in MS3 (cf. section 2.2.5). The data in fig. 2.7 show a variation of both BTE and AB1 microphones throughout the MSs. Moreover, the small time scale allows to

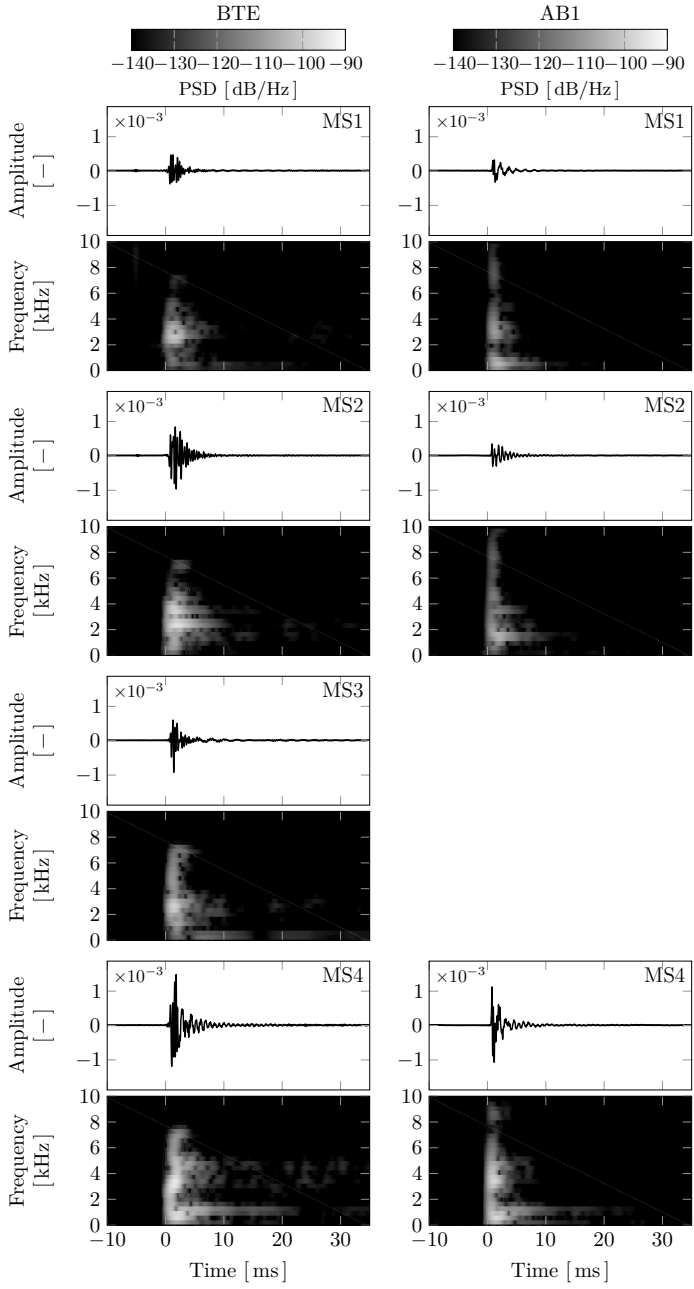


Figure 2.7: DIRs and corresponding spectrograms from BTE and AB1 microphone signals obtained with a 0.71 V_{RMS} ESS stimulus in MS1 to MS4.

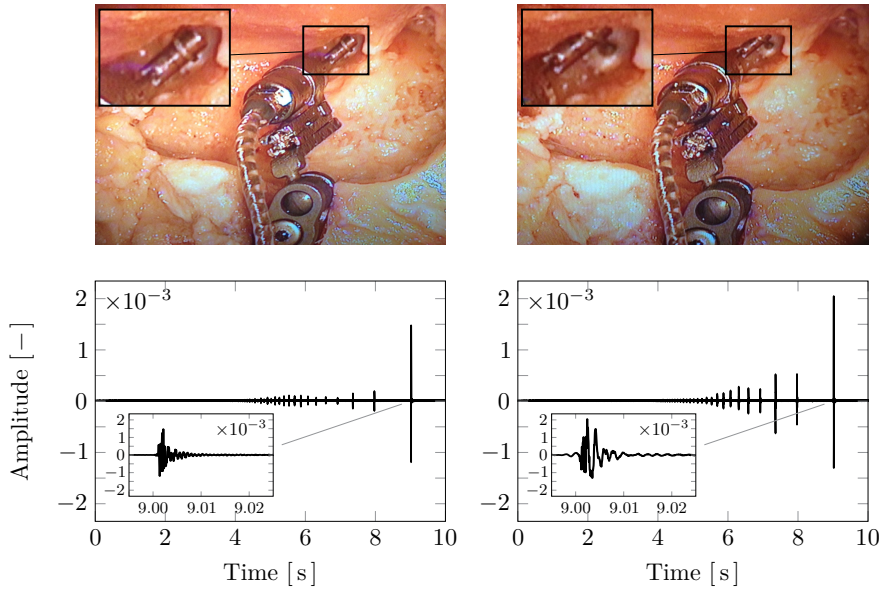


Figure 2.8: Top: Correct and incorrect placement of the Codacs actuator (top-left and top-right, respectively). The incorrectly implanted actuator touches the surrounding bone of the posterior wall of the external ear canal (compare the magnified area). Bottom: DTR from BTE microphone signal using a 0.71 V_{RMS} ESS stimulus from the corresponding surgical cases shown above; the insets show the detail of the DIR only. In each inset, the boundaries of the y-axis are kept the same as in the DTR plot, while only 30 ms of signal are shown on the x-axis.

appreciate the variation in the decay time of the responses over different MSs. In some cases, e. g. MS3 and MS4, the DIRs exhibit a slower decay in some narrow frequency bands (at approximately 500 Hz and between 800 to 900 Hz, respectively) compared to the remaining part of the spectrum, giving rise to an oscillating behavior that can last up to 20 ms. This effect can be seen in the time-domain signals but it is more easily observed in the spectrograms.

In MS4, we also investigated the variation of the nonlinear response for the case of a correctly implanted actuator to that of an incorrectly implanted one, top-left and top-right of fig. 2.8, respectively. In the same figure, the DTR is shown for both cases, with the same scaling. It can be seen how a wrong position strongly increases the nonlinearity. The DIR shows, too, an increased energy content visible from both the increased maximum amplitude and from the slower decay time (the latter can be noticed in the insets of the plots).

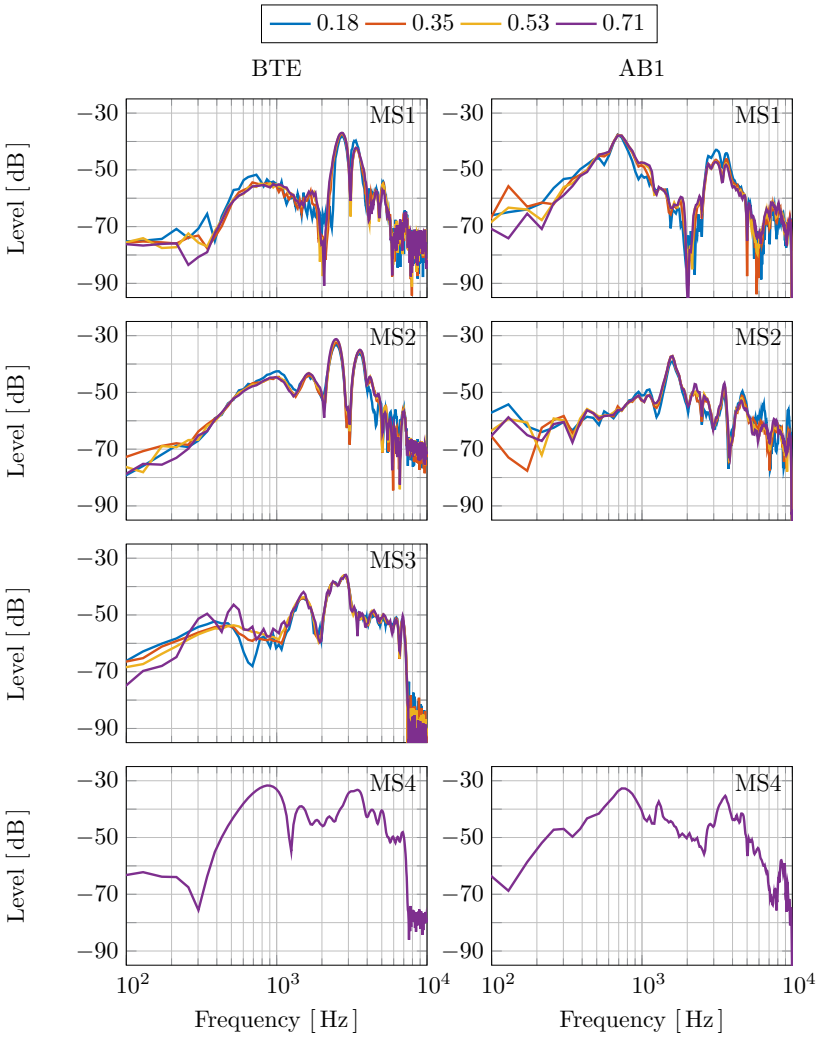


Figure 2.9: Magnitude response of the DIRs from BTE and AB1 microphone signals for the four tested levels in MS1 to MS4.

The magnitude responses of the DIRs from the BTE and AB1 microphone signals from MS1 to MS4 are shown in fig. 2.9. Each plot contains four magnitude responses, corresponding to the different ESS stimulus levels. The recordings of both BTE and AB1 microphones show variability among different MSs, e. g. the shift of the positions of peaks and dips and the change in the

scaling. The magnitude response of the DIRs in a single session, for a single microphone, do not show a good degree of similarity in the whole spectrum. Specifically, a higher variability is shown in the low frequencies and in the frequency regions where the actuator behaves nonlinearly (80 to 500 Hz). The difference between the curves corresponding to the lowest and to the highest stimuli ranges between 4 to 14 dB for the BTE recordings and between 3 to 5 dB for the AB1 recordings.

2.3.2 Application of the microphones calibration measurements

Exploiting the data from the calibration of the microphones, cf. section 2.2.6, allows the mapping of the digital levels into dB SPL, hence simplifying the within-session comparisons by reducing the effect of each microphone transfer function from the magnitude responses. Unlike the results in fig. 2.9, in fig. 2.10 we can see that an increase in the stimulus level corresponds to an increase in the measured level. The effect is more easily noticeable in those frequency bands with energy well above the averaged background noise (from 5 repetitions, i. e. reduced by approximately 7 dB and defined bg_{-7} in the figure), indicated by the black dashed line. Several differences can be seen throughout the different MSs and involve a frequency shift, as well as different magnitude, of peaks and dips in the PSDs, e. g. in MS2. By comparing the two microphone signals, mainly in MS1 and MS2, we can see that the peak centered around 700 Hz in MS1 and the peak at approximately 1500 Hz in MS2 show a roughly 10 dB stronger amplitude in the AB1 microphone PSD than the corresponding peak in the BTE microphone PSD. Other PSD components are shown to have a higher energy content in the BTE microphone signal: for instance, the two peaks at high frequency in MS1 and MS2 or the components between 200 to 1000 Hz in MS2. In MS4 there is a (visually) better correspondence, especially below 4000 Hz, between the components with highest energy in the BTE and AB1 microphone PSDs, still with some variability.

The lower panels of fig. 2.10 also show the change in the microphone PSD levels going from the correctly implanted actuator to the incorrectly implanted actuator case. The resonance peak located at approximately 750 Hz in both microphone PSDs undergoes a frequency shift to approximately 500 Hz. However, the change in the magnitude of this component is very limited in the AB1 microphone PSD while it reaches up to 15 dB in the BTE microphone PSD. Other differences can be also observed, such as the increase in the higher part of the spectrum for the AB1 microphone PSD, which is of little concern for the BTE microphone PSD due to its steep high-frequency cutoff.

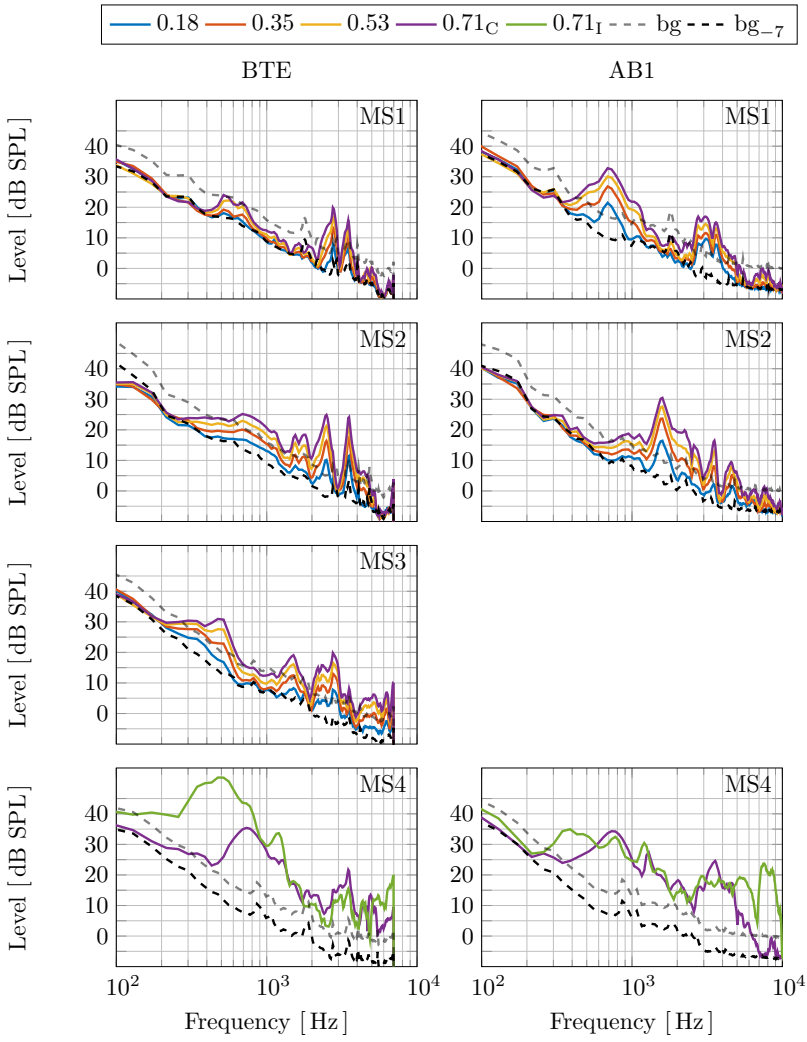


Figure 2.10: Calibrated microphone PSDs in dB SPL of the BTE and AB1 microphones signals for the four tested levels, the background noise (bg) and the averaged background noise (bg_{-7}) in MS1 to MS4. The lower panels also include a comparison between the correctly implanted actuator (0.71_C) and the incorrectly implanted actuator (0.71_I) performed in MS4.

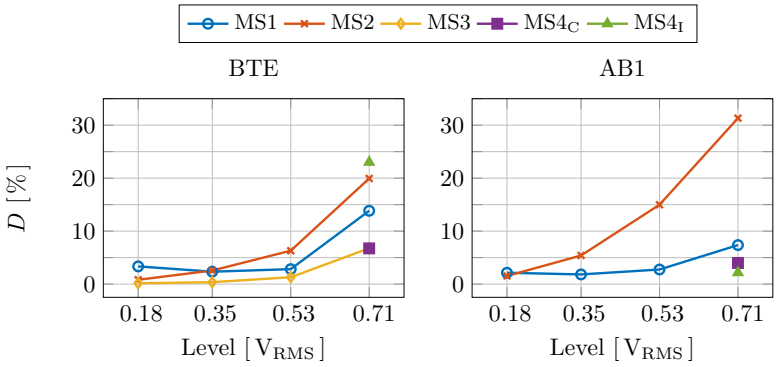


Figure 2.11: Distortion, D, calculated from BTE and AB1 microphone DTRs at different levels in MS1 to MS4 (the latter defined as $MS4_C$ in the legend), including the incorrectly implanted actuator results from MS4 ($MS4_I$).

2.3.3 Nonlinearity of the system

We conclude the section with the results of the D measure, cf. section 2.2.4, presented in fig. 2.11. The values of D are estimated from the DTRs, i. e. where no microphone calibration is applied and, hence, only a same-microphone comparison is possible. The data show that an increase in the nonlinear response can be expected when increasing the stimulus level and that the nonlinear response is subject to variability between the different sessions. For the BTE microphone, the recordings in MS3 returns the lowest D values, while those in MS2 returns the greatest D values. A strong increase, slightly lower than 20 %, of the distortion measure from the correctly implanted actuator case to the incorrectly implanted actuator case for the BTE microphone can be observed in MS4. However, such an increase is not observed for the AB1 microphone, similarly to what is shown in the lower panels of fig. 2.10.

2.4 Discussion

2.4.1 Actuator input precompensation

An important issue to be mentioned about the measurements in this paper involves the concept of input precompensation, see section 2.2.1. In particular, the difference between cadaver and patient input precompensation.

In implanted patients, a precompensation based on the intraoperatively-measured ATF is applied in the clinical software. However, in the postoperative phase, the intraoperatively-measured ATF resonance can be damped as well as slightly shifted by integration with the bony structures, and by soft tissue overgrowth due to healing and scarring, making the precompensation less effective. In the cadaver measurements we performed for this study, the precompensation was solely based on actuators ATF measured on the bench. Thus, the coupling and implantation can make, again, the precompensation less effective. Due to these limitations of the precompensation procedure, the design of a feedback canceller should be flexible enough to cope with both implantation and/or postoperative changes. Such additional flexibility can be obtained by using, e. g., an AFC algorithm [84].

2.4.2 Feedback path measurements and specimen-dependent variabilities

The results from section 2.3 suggest that the signals recorded with the same microphone in different MSs varies in both envelope and amplitude throughout the sessions. This can be seen both through the time variability of the signals in fig. 2.7 and through the frequency shift, as well as different magnitude, of peaks and dips from the magnitude responses in fig. 2.9. Due to the lack of microphone compensation in figs. 2.7 and 2.9, when visually comparing the content of different microphone signals (i. e. when performing an intermicrophone comparison), one should remember that the effects of the microphones' characteristics such as transfer function and microphone sensitivity are included in the microphone signals. Thus, the information in the figures depicting time-domain signals will be mainly used for the purpose of same-microphone comparison among MSs.

The within-session variability of the magnitude responses shown in fig. 2.9 can be accounted for by two main reasons: First, at low frequencies the SNR is the lowest, causing the estimation procedure to worsen. Second, the responses were calculated using only the linear part of the DTRs, meaning that the energy from the frequency components triggering the nonlinear behavior of the actuator has been removed. Nevertheless, a good correspondence of the magnitude responses can be seen at the resonant frequencies. This correspondence suggests that a similar behavior could be expected even at lower levels, matching more closely with the clinical range of operation. Had the experiments been conducted in a room with a higher SNR, we could reasonably expect an even lower within-session variability.

The intersession diversity seems to indicate that the changes in the specimen

head morphology can cause variability in the feedback path response. This aspect should be recalled if the data from individual measurements were to be used in the fitting procedure of the Codacs DACI. In fact, the ESS technique could be used to estimate the specific feedback path magnitude responses of an implanted patient, returning a set of data similar to those in figs. 2.7 and 2.9 which represents useful information for the design of a frequency-domain feedback canceller, a common approach employed in hearing aids applications [79, 84]. These results could, in fact, be exploited as prior knowledge to tune the parameters of a feedback canceller during the fitting procedure, such as the filter length or the overall scaling factor of the filter coefficients, to provide faster convergence to the desired solution. Additionally, the maximum of each magnitude response from the BTE results in fig. 2.9 can be used, disregarding phase information, as an estimate of the feedback margin of the closed-loop system. This means that the feedback margin in the tested specimen, at the tested stimulus levels, and for the proposed simplifying assumptions [a flat, unit-magnitude transfer function for both $G(\omega)$ and $M(\omega)$, and a perfect compensation of $A(\omega)$] ranges between 29 to 37 dB. Since, in real applications, the assumptions made for the transfer functions $A(\omega)$ and $M(\omega)$ might not always hold, and the forward path gain $G(\omega)$ can easily reach values on the order of 60 dB in some frequency bands [141], the risk of triggering instabilities exists.

Similar considerations can be made for the oscillations in fig. 2.7: although the duration of these oscillations is very limited in time for our measurements, their effect could be worsened in a closed-loop scenario if these frequency bands were to contain, at a certain time, one or more unstable frequencies of the closed-loop system.

2.4.3 Effects of an incorrect implantation and nonlinear behavior

The data from the comparison between a correctly and incorrectly implanted actuator have suggested that an incorrect implantation could be detrimental in different ways when considering system feedback (cf. figs. 2.8, 2.10 and 2.11). The frequency shift and the increased energy content in the incorrectly implanted actuator case are caused by the drastic change in the acoustic loading of the system. The direct acoustic coupling between the actuator and the bone of the posterior wall of the external ear canal causes an increase in the acoustic impedance of the system, motivating the measured differences. This might indicate a possible correlation between incorrect implantation and induced acoustic feedback problems (a signal with greater energy would result in a greater feedback), possibly being the cause of the acoustic feedback described by [48].

Furthermore, the previously-mentioned detrimental effect of the oscillations in the DIR in combination with the longer decay time resulting from an incorrect implantation could lead, more likely than in the correctly implanted case, to instability in the closed-loop system. Finally, the presence of stronger nonlinearities compared to the correctly implanted actuator case (cf. fig. 2.11) is a further indication that complications could arise from incorrect implantation of the actuator, since an increase of the nonlinearity could cause a traditional (linear) feedback canceller to fail in its feedback path estimation procedure and, hence, ultimately fail in its feedback cancellation task.

As for the data discussed in the previous section, the data shown in the lower panels of fig. 2.10 only give a qualitative description of the possible differences arising from an incorrect implantation. From the single analyzed case, it would be unreasonable trying to extrapolate a threshold on the BTE-to-AB1 level difference discriminating between correct vs. incorrect implantation. Nevertheless, if we assume that the data illustrated in the lower panels of fig. 2.10 represent a standard outcome for a correct vs. incorrect implantation test, the use of the described differences on the resonance peak, such as the frequency shift, the level increase in the BTE PSD, and the constant level in the AB1 PSD, could indeed be combined and used in a quantitative descriptor. It must be noticed that, due to the strong intermicrophone and intersession variability shown in fig. 2.10, such a descriptor would, likely, be limited to a within-subject use. A possible scenario would involve, e. g., the monitoring of an implanted patient over time: in such a case, a variation in the response over time (using a postoperative healing measurement as a reference) might indicate possible complications in the implantation.

A more quantitative and simplified description of the nonlinearities has been given by means of the distortion measure D. The data have shown that increased nonlinearities are to be expected as the stimulation level increases, possibly leading to problems when using a standard linear feedback canceller. Additionally, they have pointed out how an incorrect implantation strongly increases the value of D obtained from the BTE microphone, again making the feedback problems more severe. On the contrary, the values of D obtained from the AB1 microphone in both a correct and an incorrect implantation are comparable, suggesting that most of the additional energy from the skull vibrations does not get converted into sound radiated outside the head. The simplified description provided by the single D values, as opposed to the more dense PSDs information, combined with the seemingly (although not statistically) relevant difference of D between correctly and incorrectly implanted results from BTE and AB1 microphones, could be used to speculate the use of D as a numerical indication of an incorrect implantation. For instance, by comparing values of D measured both intra- and postoperatively, one could

assess whether some postoperative complications might have occurred. Such a possibility was not investigated in the present work, but will be considered in future investigations.

2.4.4 Difference between structure-borne mechanical and acoustic feedback

We mentioned that, in fig. 2.10, the differences between microphone signals recorded in the same session, at the same level, should be attributable to the feedback path alone. This simplification might not be met within the whole frequency spectrum in our measurements, since the noise floors of both the microphones, respectively 10 and 5 dB SPL [142] for the BTE and the AB1 microphone, are comparable with the lower recorded levels. However, above these values the comparison should hold. Additionally, various spectral portions of the measured signals were in proximity of the noise floor and this partially limits the usability of the microphone signal PSDs within those frequencies. Fortunately, this is not necessarily the case for the DIRs in fig. 2.9, given the good performance of the ESS technique in the presence of pink background noise, similar to the one recorded during the measurements [137, 138].

The overall trend in the results seems to confirm our hypothesis of a difference in the feedback paths defined by the two microphones and it is plausible to attribute part of the discrepancy to the presence of a structure-borne mechanical component in the feedback path to the BTE microphone not being picked-up by the AB1 microphone. Specifically, we believe that the direct mechanical coupling of the sound processor (and thus the BTE microphone) to the implant (through the bone and soft tissue layers) is responsible for this phenomenon. For this reason, we were initially expecting stronger peak components in the BTE microphone PSD compared to the AB1 microphone PSD, such as in the extreme case of an incorrect implantation in the lower panels of fig. 2.10. However, in the comparison between the two microphone signals from MS1 and MS2 in fig. 2.10 the opposite occurs. The higher value of the AB1 microphone PSDs could be due to the specific position of the two microphones with respect to the actuator and the tested ear. Given the proximity between the microphones and the sound radiating source, near-field effects can be expected, such as strongly directional radiation patterns possibly leading to the different levels measured by the microphones. Nevertheless, we cannot necessarily rule out the possibility of a measurement artifact. Almost no level discrepancy was shown between the BTE and AB1 microphone PSDs in MS4. This could be due to the fact that the previously mentioned effects are dependent on the specific anatomy of the specimen and, hence, not expected in each measurements. Although these are just hypotheses accounting for said difference and the true mechanism was not

investigated further, low values of the BTE microphone PSDs are preferred in a real scenario, as they would limit the occurrence of acoustic feedback.

2.5 Conclusions

In this study, we have measured the feedback path of the Cochlear Codacs DACI on fresh frozen cadaver heads. The measurements provide four important answers to the same number of initial general objectives: 1) The difference between the signals recorded by the different microphones in the different MSs indicate that there is clearly a dependence of the measured feedback path on the implanted specimen; 2) An incorrect implantation can strongly affect the measured feedback; 3) The response of the actuator is indeed nonlinear at the measured levels and the nonlinearities characterizing such response are shaped by the specimen-dependent mechanical coupling; 4) The difference between the feedback paths from the actuator to the different microphones points out the presence of a structure-borne mechanical feedback component in the BTE microphone, more slowly decaying than the acoustic component recorded in both microphones. Combining the outcomes of these findings, especially from 2) and 3), in a closed-loop scenario could exacerbate the different individual problems, potentially leading to acoustic feedback artifacts.

The next step of our investigation will focus on the testing of the DACI in a more realistic closed-loop scenario, to observe whether the use of some simple feedback-cancellation strategies can cope with possible feedback problems and to evaluate whether the information described in this paper can be effectively used to improve the cancellation performance of an AFC algorithm.

Acknowledgments

The authors would like to thank Dr. A. T. Rosell (Technical University of Denmark, DK), Dr. J. Walraevens and A. Gozzi (Cochlear Technology Centre Belgium, BE), and Dr. J.-M. Gérard (Saint-Luc University Hospital, BE).

Chapter 3

The early prototype of a bone conduction implant concept

Chapter 3 cannot be shared publicly due to confidentiality obligations.

Part II

Adaptive Feedback Cancellation Algorithms

Chapter 4

The frequency-domain PEM-based adaptive filter

An All-Frequency-Domain Adaptive Filter with PEM-Based Decorrelation for Acoustic Feedback Control

Giuliano Bernardi, Toon van Waterschoot, Jan Wouters, and Marc Moonen

Published in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA '15)*, Oct. 2015, pp. 1–5, DOI: 10.1109/WASPAA.2015.7336931.

© 2015 IEEE. Reprinted, with permission, from:
G. Bernardi, T. van Waterschoot, J. Wouters, and M. Moonen, "An all-frequency-domain adaptive filter with PEM-based decorrelation for acoustic feedback control," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA '15)*, (New Paltz, NY), pp. 1–5, Oct. 2015.
Changes include layout, representation, and minor editing aspects.

The candidate's contributions as first author include: literature study, co-development of the presented algorithms, co-design of the evaluation experiments, software implementation and computer simulations, co-formulation of the conclusions, text redaction and editing.

Abstract

A new method for acoustic feedback control is derived, combining an adaptive filter with a decorrelation strategy based on the prediction-error method (PEM) entirely in the frequency domain. This all-frequency-domain method extends existing methods combining a frequency-domain adaptive filter (FDAF) and time-domain PEM-based decorrelation. Such a combined approach is critical, since a standard adaptive feedback cancellation (AFC) procedure without proper decorrelation yields biased estimation results and hence does not provide suitable acoustic feedback control. The proposed modification allows the reduction of the computational complexity of the algorithm, while maintaining comparable performance.

4.1 Introduction

Frequency-domain adaptive filter (FDAF) implementations have been extensively used in digital signal processing, since their introduction more than two decades ago, given the advantages they offer compared to classical time-domain implementations. Not only can a frequency-domain implementation lead to a reduction in the computational complexity of the algorithm, working in the frequency-domain also allows the exploitation of the decorrelation properties of the discrete Fourier transform to achieve improved performance [113, 84, 143].

Several applications employing a FDAF implementation require additional operations to provide the assumed working conditions for the adaptive filter. An example of such additional operations is a decorrelation, necessary to reduce the correlation between the adaptive filter desired signal and undesired interference, inherent in a closed-loop system identification scenario. A whitening prefilter based on the prediction-error method (PEM) can be successfully applied for decorrelation [108].

Acoustic feedback control is an application example commonly addressed using adaptive feedback cancellation (AFC), which indeed requires an additional decorrelation. This can be understood referring to fig. 4.1, by noticing that the desired signal $v(t)$ and the undesired interference $x(t)$ are related by a closed-loop transfer function (TF) [46]. The closed-loop TF depends on the different TFs characterizing the system, depicted as blocks in fig. 4.1: $F(q,t)$ and $\hat{F}(q,t)$ represent the true and estimated feedback path, respectively, and $G(q,t)$ represents the forward path. All the TFs are defined by means of the delay operator q^{-1}, i. e. $q^{-k}u(t) = u(t-k)$.

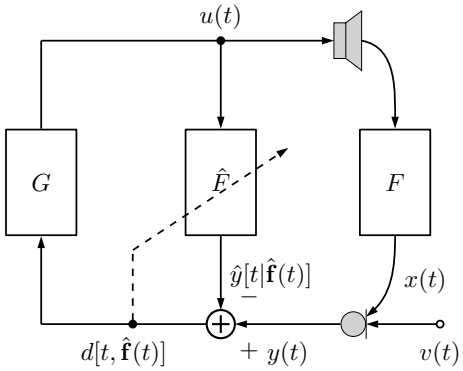


Figure 4.1: General AFC scenario.

The use of a standard adaptive filter in this scenario leads to a biased estimate $\hat{F}(q,t)$, due to the correlation between $v(t)$ and $x(t)$ [97]. The problem of a biased estimate is exacerbated by the presence of the closed-loop system, in which a poor estimation can also induce unstable behavior such as howling. Therefore, a procedure to effectively decorrelate $v(t)$ and $x(t)$ must be introduced.

Different methods combining the use of an adaptive filter and a prediction-error method (PEM)-based prefilter for the decorrelation have been discussed in several papers: the earliest such method relies on time-domain implementations for both the adaptive filter and the PEM-based decorrelation [97, 93, 83]. A time-domain implementation for the decorrelation combined with a FDAF, referred to as PEM-based frequency-domain adaptive filter (PEM-FDAF), has also been proposed [84, 144].

In this paper, we present a modification of the existing methods providing a frequency-domain implementation of both the adaptive filter and the PEM-based decorrelation. This all-frequency-domain method reduces the computational complexity of the PEM-FDAF while maintaining comparable performance. Moreover, by means of some simplifications, the new method allows to lump all the decorrelation operations into the adaptive filter stepsize, possibly leading to a new interesting interpretation of the decorrelation role.

In section 4.2, we introduce the existing PEM-FDAF method with time-domain PEM-based decorrelation. In section 4.3, we derive the all-frequency-domain method relying on frequency-domain PEM-based decorrelation and, in section 4.4, we provide a computational complexity comparison. In section 4.5, we show some simulation results highlighting the similarity between existing

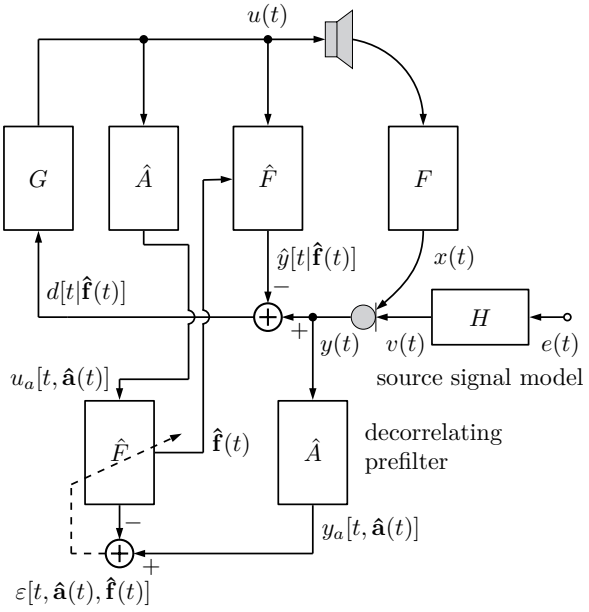


Figure 4.2: Block diagram of an AFC including PEM-based decorrelation.

and proposed methods, in terms of convergence, added stability and sound quality. Finally, the conclusions are drawn in section 4.6.

4.2 PEM-based frequency-domain adaptive filter

The PEM-FDAF method [84, 144] for AFC can be derived from the standard FDAF, by introducing a set of prefiltering operations, as illustrated in fig. 4.2. The microphone signal $y(t)$ represents the sum of the desired source signal $v(t)$ and the undesired feedback signal $x(t)$. The true acoustic feedback path $\mathbf{f}(t)$ is assumed to be of known, finite order n_F and is estimated by the adaptive filter $\hat{\mathbf{f}}(t)$, of order $n_{\hat{F}} = n_F$; the estimate $\hat{\mathbf{f}}(t)$ is then used to calculate $\hat{y}[t|\hat{\mathbf{f}}(t)]$, i. e. a prediction of the undesired signal $x(t)$, in order to remove the contribution of $x(t)$ from the microphone signal $y(t)$.

The additional blocks in fig. 4.2, compared to fig. 4.1, define the time-varying autoregressive (AR) source signal model $H(q,t) = \frac{1}{A(q,t)}$ and the estimated inverse source signal model $\hat{A}(q,t)$, both of order n_H. By means of $H(q,t)$, the source signal $v(t)$ can be modeled using the white noise sequence $e(t)$ as input

signal, i. e. $v(t) = H(q,t)e(t)$. The modeling choice on $H(q,t)$ is done since a low-order AR model has been shown to provide a good approximation for speech signals [145, 99]. The role of $\hat{A}(q,t)$, also expressed through the corresponding time-domain parameter vector $\hat{\mathbf{a}}(t)$, is to prefilter the microphone, loudspeaker and hence the error signal, obtaining $y_a[t, \hat{\mathbf{a}}(t)]$, $u_a[t, \hat{\mathbf{a}}(t)]$ and $\varepsilon[t, \hat{\mathbf{a}}(t), \hat{\mathbf{f}}(t)]$, respectively, used to perform the adaptation procedure. In the following, the dependencies on $\hat{\mathbf{a}}(t)$ and $\hat{\mathbf{f}}(t)$ will be omitted.

To describe the PEM-FDAF algorithm, we must first introduce the discrete time frame and frequency frame variables, k and l, respectively. The M-samples loudspeaker signal frame at (frame) time $k \in \mathbb{Z}$ and its prefiltered version are defined as follows:

$$\mathbf{u}(k) = [u(kL - M + 1) \quad \dots \quad u(kL)]^{\mathrm{T}} \tag{4.1a}$$

$$\mathbf{u}_a(k) = [u_a(kL - M + 1) \quad \dots \quad u_a(kL)]^{\mathrm{T}}, \tag{4.1b}$$

with L denoting the frame shift.

The microphone signal frame and its prefiltered version are constrained to R-samples — typically $R = L = M/2$ and $R = n_{\hat{F}}$ [113, 144] — to accommodate for the overlap-and-save procedure:

$$\mathbf{y}(k) = [y(kL - R + 1) \quad \dots \quad y(kL)]^{\mathrm{T}} \tag{4.2a}$$

$$\mathbf{y}_a(k) = [y_a(kL - R + 1) \quad \dots \quad y_a(kL)]^{\mathrm{T}}. \tag{4.2b}$$

For the loudspeaker and microphone signal, the prefiltering can be expressed as follows:

$$u_a(kL - j) = \hat{A}(q,k)u(kL - j) \quad j = 0, \dots, M - 1 \tag{4.3a}$$

$$y_a(kL - j) = \hat{A}(q,k)y(kL - j) \quad j = 0, \dots, R - 1. \tag{4.3b}$$

The frequency-domain versions of the loudspeaker signal frame, before and after prefiltering, can be expressed introducing the discrete Fourier transform matrix $\mathbf{F}_M$ of size $M \times M$:

$$\mathbf{U}(k) = \operatorname{diag}\{\mathbf{F}_M \, \mathbf{u}(k)\} \tag{4.4a}$$

$$\mathbf{U}_a(k) = \operatorname{diag}\{\mathbf{F}_M \, \mathbf{u}_a(k)\}. \tag{4.4b}$$

The error signal frame is defined as the difference between the microphone signal frame and the estimated feedback signal frame, where the latter is calculated

in the frequency domain by circular convolution, linearized via the projection matrix $\mathbf{Q} = [\mathbf{0}_R \quad \mathbf{I}_R]^{\mathrm{T}}$:

$$\mathbf{d}(k) = \mathbf{y}(k) - \hat{\mathbf{y}}(k) \tag{4.5a}$$

$$\hat{\mathbf{y}}(k) = \mathbf{Q}^{\mathrm{T}}\mathbf{F}_M^{-1}\,\mathbf{U}(k)\hat{\mathbf{F}}(k), \tag{4.5b}$$

with $\hat{\mathbf{F}}(k) = \mathbf{F}_M \begin{bmatrix}\hat{\mathbf{f}}^{\mathrm{T}}(k) & \mathbf{0}_{1\times R}\end{bmatrix}^{\mathrm{T}}$.

The AR coefficients $\hat{\mathbf{a}}(k) = [1 \quad \hat{a}_1(k) \quad \dots \quad \hat{a}_{n_H}(k)]^{\mathrm{T}}$ estimate the denominator coefficients of $H(q,k)$ and, equivalently, define the numerator coefficients of $\hat{A}(q,k)$. These coefficients are estimated using the Levinson-Durbin algorithm [90, pp. 254-264] on the concatenation of the error signal frames $\mathbf{d}(k)$ and $\mathbf{d}(k-1)$, and are used to calculate the prefiltered quantities $\mathbf{u}_a(k)$ and $\mathbf{y}_a(k)$, cf. (4.3), (4.1b) and (4.2b). The prediction error signal frame $\boldsymbol{\varepsilon}(k)$ is, in turn, obtained by means of these prefiltered quantities:

$$\boldsymbol{\varepsilon}(k) = \mathbf{y}_a(k) - \hat{\mathbf{y}}_a(k) \tag{4.6a}$$

$$\hat{\mathbf{y}}_a(k) = \mathbf{Q}^{\mathrm{T}}\mathbf{F}_M^{-1}\,\mathbf{U}_a(k)\hat{\mathbf{F}}(k). \tag{4.6b}$$

Finally, the prediction error signal frame is transformed to the frequency domain and used, together with the frequency-domain prefiltered loudspeaker signal frame, in the stochastic gradient iteration to update the estimated $\hat{\mathbf{F}}(k)$:

$$\boldsymbol{\mathcal{E}}(k) = \mathbf{F}_M\,\mathbf{Q}\boldsymbol{\varepsilon}(k) \tag{4.7a}$$

$$\Delta\hat{\mathbf{F}}(k) = \mathrm{diag}\{\boldsymbol{\mu}(k)\}\mathbf{U}_a^{\mathrm{H}}(k)\boldsymbol{\mathcal{E}}(k) \tag{4.7b}$$

$$\hat{\mathbf{F}}(k+1) = \hat{\mathbf{F}}(k) + \mathbf{F}_M\,\left[\mathbf{I}_M - \mathbf{Q}\mathbf{Q}^{\mathrm{T}}\right]\mathbf{F}_M^{-1}\,\Delta\hat{\mathbf{F}}(k). \tag{4.7c}$$

The stepsize $\boldsymbol{\mu}(k)$ normalizes $\Delta\hat{\mathbf{F}}(k)$ using the instantaneous power of the adaptive procedure's reference signal, and its l-th element at time k is defined as follows:

$$\mu(k,l) = \mu_{\mathrm{fix}}\left[\alpha + |U_a(k,l)|^2\right]^{-1}, \tag{4.8}$$

where μ_{fix} is a real, constant, and frequency-independent value. A summary of the PEM-FDAF algorithm is given in algorithm 1.

4.3 Frequency-domain PEM-based adaptive filter

The proposed all-frequency-domain method, referred to as frequency-domain prediction-error-method-based adaptive filter (FD-PEMAF), modifies the PEM-

Algorithm 1 PEM-based Frequency-domain adaptive filter (PEM-FDAF) [144].

Initialize $R = L = M/2$, $\hat{\mathbf{F}}(0) = \mathbf{d}(0) = \mathbf{0}_{M\times 1}$
for $k = 1, 2, \ldots$ **do**
  $\mathbf{U}(k) = \text{diag}\left\{\mathbf{F}_M \left[u(kL - M + 1), \ldots, u(kL)\right]^{\text{T}}\right\}$
  $\hat{\mathbf{y}}(k) = \mathbf{Q}^{\text{T}}\mathbf{F}_M^{-1}\,\mathbf{U}(k)\hat{\mathbf{F}}(k)$
  $\mathbf{d}(k) = \mathbf{y}(k) - \hat{\mathbf{y}}(k)$
  $\hat{\mathbf{a}}(k) = \texttt{Levinson} - \texttt{Durbin}\left([\mathbf{d}(k); \mathbf{d}(k-1)]\right)$
  $u_a(kL - j) = \hat{A}(q, k)u(kL - j), \quad j = 0, \ldots, M - 1$
  $y_a(kL - j) = \hat{A}(q, k)y(kL - j), \quad j = 0, \ldots, R - 1$
  $\mathbf{U}_a(k) = \text{diag}\left\{\mathbf{F}_M\,\mathbf{u}_a(k)\right\}$
  $\hat{\mathbf{y}}_a(k) = \mathbf{Q}^{\text{T}}\mathbf{F}_M^{-1}\,\mathbf{U}_a(k)\hat{\mathbf{F}}(k)$
  $\boldsymbol{\varepsilon}(k) = \mathbf{y}_a(k) - \hat{\mathbf{y}}_a(k)$
  $\boldsymbol{\mathcal{E}}(k) = \mathbf{F}_M\,\mathbf{Q}\boldsymbol{\varepsilon}(k)$
  $\mu(k, l) = \mu_{\text{fix}}\left[\alpha + |U_a(k, l)|^2\right]^{-1}$
  $\Delta\hat{\mathbf{F}}(k) = \text{diag}\{\boldsymbol{\mu}(k)\}\mathbf{U}_a^{\text{H}}(k)\boldsymbol{\mathcal{E}}(k)$
  $\hat{\mathbf{F}}(k + 1) = \hat{\mathbf{F}}(k) + \mathbf{F}_M\left[\mathbf{I}_M - \mathbf{Q}\mathbf{Q}^{\text{T}}\right]\mathbf{F}_M{}^{-1}\Delta\hat{\mathbf{F}}(k)$
end for

FDAF by performing the prefiltering operations in the frequency-domain. The prefilter is still estimated from the time-domain error signal, but the adaptive filter calculation is now fully performed in the frequency domain. In order to implement the prefiltering operations correctly, the frame shift is reduced by a number of samples equal to the prefilter length, i. e. from R to $L = R - n_H$. The vector of AR coefficients $\hat{\mathbf{a}}(k)$ is estimated as in the PEM-FDAF, and subsequently transformed to the frequency domain:

$$\hat{\mathbf{A}}(k) = \text{diag}\left\{\mathbf{F}_M \left[\hat{\mathbf{a}}^{\text{T}}(k) \quad \mathbf{0}_{1\times(M-n_H-1)}\right]^{\text{T}}\right\}. \tag{4.9}$$

The L-samples prediction error signal frame $\boldsymbol{\varepsilon}(k)$ is defined by means of (4.9), and transformed to the frequency domain as follows:

$$\boldsymbol{\varepsilon}(k) = \mathbf{Q}_{n_H}^{\text{T}}\mathbf{F}_M^{-1}\,\hat{\mathbf{A}}(k)[\mathbf{Y}(k) - \mathbf{U}(k)\hat{\mathbf{F}}(k)] \tag{4.10a}$$

$$\boldsymbol{\mathcal{E}}(k) = \mathbf{F}_M\,\mathbf{Q}_{n_H}\boldsymbol{\varepsilon}(k). \tag{4.10b}$$

(4.10a) includes the frequency-domain version of the microphone signal frame $\mathbf{Y}(k) = \mathbf{F}_M\,\mathbf{Q}\mathbf{y}(k)$, and provides the concurrent frequency-domain filtering of the microphone signal frame $\mathbf{y}(k)$, via the prefilter $\hat{\mathbf{a}}(k)$, and of the loudspeaker

Algorithm 2 Constrained and unconstrained frequency-domain PEM-based adaptive filter (FD-PEMAF and uFD-PEMAF, respectively).

Initialize $M = 2R$, $L = R - n_H$, $\hat{\mathbf{F}}(0) = \mathbf{d}(0) = \mathbf{0}_{M\times 1}$
for $k = 1, 2, \ldots$ **do**
  $\mathbf{U}(k) = \text{diag}\left\{\mathbf{F}_M \left[u(kL - M + 1), \ldots, u(kL)\right]^{\text{T}}\right\}$
  $\mathbf{Y}(k) = \mathbf{F}_M \mathbf{Q}\mathbf{y}(k)$
  $\mathbf{d}(k) = \mathbf{y}(k) - \mathbf{Q}^{\text{T}}\mathbf{F}_M^{-1}\mathbf{U}(k)\hat{\mathbf{F}}(k)$
  $\hat{\mathbf{a}}(k) = \texttt{Levinson} - \texttt{Durbin}\left([\mathbf{d}(k); \mathbf{d}(k-1)]\right)$
  $\hat{\mathbf{A}}(k) = \text{diag}\left\{\mathbf{F}_M \left[\hat{\mathbf{a}}^{\text{T}}(k) \quad \mathbf{0}_{1\times(M-n_H-1)}\right]^{\text{T}}\right\}$
  if unconstrained **then**
    $\mathbf{G}_{n_H} \approx L/M \cdot \mathbf{I}_M$
  else
    $\mathbf{G}_{n_H} = \mathbf{F}_M \mathbf{Q}_{n_H}\mathbf{Q}_{n_H}^{\text{T}}\mathbf{F}_M^{-1}$
  end if
  $\boldsymbol{\mathcal{E}}(k) = \mathbf{G}_{n_H}\hat{\mathbf{A}}(k)[\mathbf{Y}(k) - \mathbf{U}(k)\hat{\mathbf{F}}(k)]$
  $\mu(k,l) = \mu_{\text{fix}}\left[\alpha + |U(k,l)\hat{A}(k,l)|^2\right]^{-1}$
  $\Delta\hat{\mathbf{F}}(k) = \text{diag}\{\boldsymbol{\mu}(k)\}\hat{\mathbf{A}}^{\text{H}}(k)\mathbf{U}^{\text{H}}(k)\boldsymbol{\mathcal{E}}(k)$
  $\hat{\mathbf{F}}(k+1) = \hat{\mathbf{F}}(k) + \mathbf{F}_M \left[\mathbf{I}_M - \mathbf{Q}\mathbf{Q}^{\text{T}}\right]\mathbf{F}_M{}^{-1}\Delta\hat{\mathbf{F}}(k)$
end for

signal frame $\mathbf{u}(k)$, via the prefilter $\hat{\mathbf{a}}(k)$ and the feedback path $\hat{\mathbf{f}}(k)$. As the prediction error signal frame $\boldsymbol{\varepsilon}(k)$ is n_H-samples shorter than the error signal $\mathbf{d}(k)$, $\mathbf{Q} \neq \mathbf{Q}_{n_H} = \begin{pmatrix}\mathbf{0}_{L\times R+n_H} & \mathbf{I}_L\end{pmatrix}^{\text{T}}$.

The definition of $\Delta\hat{\mathbf{F}}(k)$, cf. (4.7b), will then become:

$$\Delta\hat{\mathbf{F}}(k) = \text{diag}\{\boldsymbol{\mu}(k)\}\hat{\mathbf{A}}^{\text{H}}(k)\mathbf{U}^{\text{H}}(k)\boldsymbol{\mathcal{E}}(k). \tag{4.11}$$

This new definition does not require the explicit calculation of the frequency-domain prefiltered loudspeaker signal $\mathbf{U}_a(k)$, thus leading to a slightly modified definition of the adaptive stepsize $\mu(k,l)$, cf. algorithm 1 and 2.

A different definition of $\Delta\hat{\mathbf{F}}(k)$ can be obtained by approximating the linearization constraint $\mathbf{G}_{n_H} = \mathbf{F}_M \mathbf{Q}_{n_H}\mathbf{Q}_{n_H}^{\text{T}}\mathbf{F}_M^{-1}$ as $\mathbf{G}_{n_H} \approx L/M \cdot \mathbf{I}_M$ in the calculation of $\boldsymbol{\mathcal{E}}(k)$, as suggested by Benesty *et al.* [146, pp. 168-170], leading to

$$\Delta\hat{\mathbf{F}}(k) = L/M \cdot \text{diag}\{\boldsymbol{\mu}(k)\}\hat{\mathbf{A}}(k)\hat{\mathbf{A}}^{\text{H}}(k)\mathbf{U}^{\text{H}}(k)\mathbf{D}(k), \tag{4.12}$$

where the order of the multiplications has been changed, due to the diagonality of both $\hat{\mathbf{A}}(k)$ and $\mathbf{U}(k)$, and $\mathbf{D}(k) = \mathbf{F}_M \, \mathbf{Qd}(k)$. Equations (4.11) and (4.12) are not equivalent, and the substitution of (4.11) with (4.12) results in a modified algorithm referred to as unconstrained FD-PEMAF (uFD-PEMAF). The motivation to choose an unconstrained version could arise, for instance, if complexity is an issue [113], since dropping the constraints allows to save two FFT operations per iteration.

Equation (4.12) allows to lump all the prefiltering operations into the stepsize, obtaining:

$$\mathrm{diag}\{\boldsymbol{\mu}_a(k)\} = L/M \cdot \mathrm{diag}\{\boldsymbol{\mu}(k)\}\hat{\mathbf{A}}(k)\hat{\mathbf{A}}^{\mathrm{H}}(k) \tag{4.13a}$$

$$\Delta\hat{\mathbf{F}}(k) = \mathrm{diag}\{\boldsymbol{\mu}_a(k)\}\mathbf{U}^{\mathrm{H}}(k)\mathbf{D}(k). \tag{4.13b}$$

A summary of the FD-PEMAF is given in algorithm 2.

In section 4.1, we mentioned that prefiltering is introduced to reduce the bias of the feedback path estimate by providing some degree of decorrelation between the loudspeaker and microphone signal. Therefore, lumping the prefilter contribution into the stepsize extends the usual set of measures affected by stepsize tuning, namely convergence speed, stability and steady-state error [90], by the estimation bias. Such a representation thus provides an interesting insight into the role of the stepsize. However, unlike other unconstrained FDAF algorithms [113], our results will show a strong performance decrease, making it a rather unappealing choice.

4.4 Complexity

In this section, we provide a complexity comparison between existing and proposed algorithms, focusing on the number of per-sample real multiplications. Assuming equal complexity for a real multiplication and a real division, a complexity of $M \log_2(M)$ for each FFT/IFFT operation [113], a real, frequency-dependent stepsize, and a total cost of $n_H^2 + (5 + M)n_H + M$ for the Levinsion-Durbin calculations [90], the three algorithms have the computational complexities shown in table 4.1. The FD-PEMAF requires one additional FFT operation compared to the PEM-FDAF, as seen in the first factor at numerator in the complexity calculations, but limits the overall complexity by removing the time-domain prefiltering operations.

Using the values chosen in the simulations discussed in the following section, i. e. $M = 160$, $R = 80$, $n_H = 15$ and $L = 65$, the cost for each new processed sample is 278, 240 and 204 for PEM-FDAF, FD-PEMAF and uFD-PEMAF,

Table 4.1: Per-sample computational complexity of the algorithms.

Algorithm	Real multiplications
PEM-FDAF	$\frac{7M\log_2(M)+19M+R+3n_H^2+n_H(7+2M+R)}{R-n_H}$
FD-PEMAF	$\frac{8M\log_2(M)+22M+n_H^2+n_H(5+M)}{R-n_H}$
uFD-PEMAF	$\frac{6M\log_2(M)+22M+n_H^2+n_H(5+M)}{R-n_H}$

respectively. This corresponds to a complexity decrease of 14 % for FD-PEMAF and 27 % for uFD-PEMAF, when compared with PEM-FDAF.

4.5 Simulation Results

The performance of the algorithms is assessed by means of three standard measures evaluating the estimation error, the achievable amplification and the sound quality. The estimation error is evaluated using the misadjustment (Mis), i. e. the normalized distance between the true and estimated feedback path [46]; the achievable amplification is assessed via the added stable gain (ASG), i. e. the increase of the maximum stable gain (MSG) due to the inclusion of the adaptive filter, assuming a spectrally flat $G(q,t)$ [84]; the sound quality is evaluated using the so-called frequency-weighted log-spectral signal distortion (SD), providing good correlation with subjective evaluations in feedback cancellation tasks [121].

Clean speech, sampled at $f_\mathrm{s} = 16\,\mathrm{kHz}$ was chosen as the source signal $v(k)$, while the feedback path $F(q,t)$ was measured in a commercial hearing aid [84]. The TF $G(q,t)$ was characterized by a delay of 65 samples and two broadband gain values: the first, referred to as K_1, corresponded to 3 dB below the MSG obtained without the inclusion of $\hat{F}(q,t)$; the second, referred to as K_2, was set to $K_2 = K_1 + 10\,\mathrm{dB}$. The estimated AR model for $v(t)$ was fixed to 15, similar to the usual low-order values used in literature [99, 70], and the frame parameters were chosen to be $M = 160$, $R = 80$ and $L = 65$ (for the PEM-FDAF too), as previously mentioned. Finally, the fixed step size and the regularization constant were set to $\mu_\mathrm{fix} = 0.01$ and $\alpha = 5 \times 10^{-5}$, respectively.

The first set of results for Mis and ASG refers to the K_1 condition and is shown using dotted lines in fig. 4.3. After 10 s, both PEM-FDAF and FD-PEMAF reach some $-8\,\mathrm{dB}$ of Mis, roughly 6 dB lower than uFD-PEMAF; similar results, with the opposite sign, exists for the ASG results (8 dB for PEM-

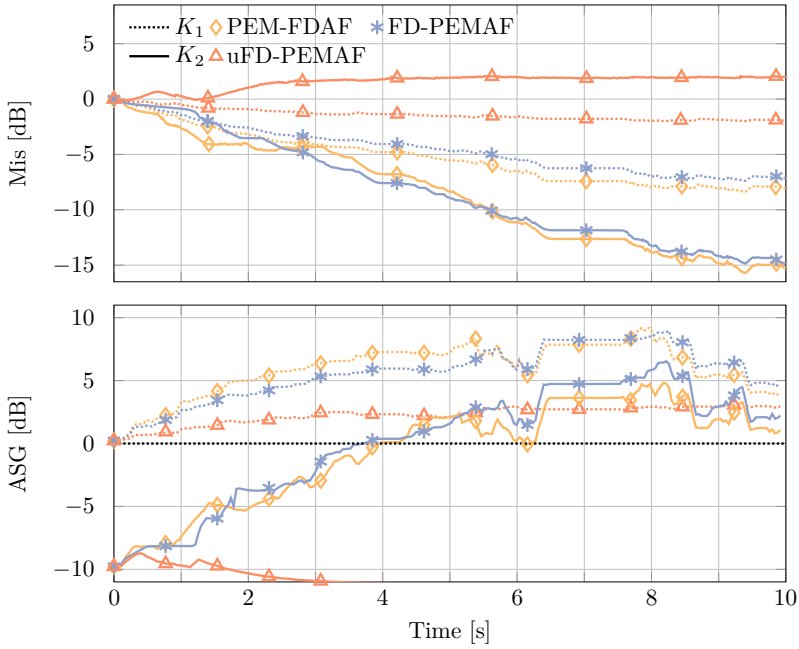


Figure 4.3: Mis and ASG performance for the three algorithms using the two tested gains.

FDAF and FD-PEMAF, and 2 dB uFD-PEMAF) after 8 s, but the constrained versions undergo a performance deterioration toward the end of the simulation, possibly due to the specific signal characteristics in such time interval. The performance mismatch between PEM-FDAF and FD-PEMAF can be attributed to the different calculation of $\mu(k,l)$, cf. algorithm 1 and 2.

The second set of results for Mis and ASG refers to the K_2 condition and is shown using solid lines in fig. 4.3. In this case, the Mis values after 10 s for PEM-FDAF and FD-PEMAF are both shifted by roughly −7 dB, while uFD-PEMAF has now convergence problems. The ASG results after 10 s for PEM-FDAF and FD-PEMAF are approximately 3 dB lower than in the K_1 condition, but both algorithms need almost 4 s to reach a greater-than-zero ASG. On the contrary, uFD-PEMAF remains heavily unstable (ASG < -8 dB) throughout the whole simulation.

Table 4.2 lists the sound quality results, by means of mean and maximum SD. Similar results, in both tested conditions, are obtained for PEM-FDAF and FD-PEMAF, while uFD-PEMAF shows a worse performance, severely worsened in the K_2 condition.

The simulation results presented in this section suggest that comparable

Table 4.2: Mean and maximum SD values for the different algorithms in the two tested scenarios.

	mean(SD)		max(SD)	
	K_1	K_2	K_1	K_2
PEM-FDAF	1.52	1.83	4.16	4.38
FD-PEMAF	1.74	2.25	4.21	4.49
uFD-PEMAF	2.79	41.96	5.23	67.53

performance is expected with PEM-FDAF and FD-PEMAF, whereas a stronger degradation can be expected using uFD-PEMAF, making the latter a highly discouraged choice.

4.6 Conclusions

An all-frequency-domain method for AFC has been derived, allowing to reduce the computational complexity of the PEM-FDAF while maintaining comparable performance in terms of estimation error, achievable amplification and sound-quality performance.

An additional simplification obtained by dropping a linearization constraint and allowing to interpret the stepsize multiplication of the adaptive algorithm as a decorrelation operation, has been shown to dramatically worsen the reference performance, especially for higher values of the forward path gain. Therefore, such a simplification is highly discouraged.

Chapter 5

The PEM-based frequency-domain Kalman filter

Adaptive Feedback Cancellation Using a Partitioned-Block Frequency-Domain Kalman Filter Approach with PEM-Based Signal Prewhitening

Giuliano Bernardi, Toon van Waterschoot, Jan Wouters, and Marc Moonen

Published in *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 9, pp. 1480–1494, Sep. 2017, DOI: 10.1109/TASLP.2017.2716188.

© 2017 IEEE. Reprinted, with permission, from:
G. Bernardi, T. van Waterschoot, J. Wouters, and M. Moonen, "Adaptive feedback cancellation using a partitioned-block frequency-domain Kalman filter approach with PEM-based signal prewhitening," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, pp. 1480–1494, Sep. 2017.
Changes include layout, representation, and minor editing aspects.

The candidate's contributions as first author include: literature study, co-development of the presented algorithms, co-derivation of the identifiability conditions, co-design of evaluation experiments, software implementation and computer simulations, co-formulation of conclusions, text redaction and editing.

Abstract

Adaptive filtering based feedback cancellation is a widespread approach to acoustic feedback control. However, traditional adaptive filtering algorithms have to be modified in order to work satisfactorily in a closed-loop scenario. In particular, the undesired signal correlation between the loudspeaker signal and the source signal in a closed-loop scenario is one of the major problems to address when using adaptive filters for feedback cancellation. Slow convergence speed and limited tracking capabilities are other important limitations to be considered. Additionally, computationally expensive algorithms as well as long delays should be avoided, for instance, in hearing aid (HA) applications, because of power constraints, important to extend battery life, and real-time implementations requirements, respectively. We present an algorithm combining good decorrelation properties, by means of the prediction-error method (PEM)-based signal prewhitening, fast convergence, good tracking behavior, and low computational complexity by means of the frequency-domain Kalman filter (FDKF), and low delay by means of a partitioned block (PB) implementation.

5.1 Introduction

Acoustic feedback control is of critical importance in several systems dealing with acoustic signals, such as public address (PA) systems and hearing aids (HAs). A lack of acoustic feedback control can lead to system instabilities causing annoying artifacts and sound degradation. Mainly in the last three decades, several methods have been developed to cope with the problem of acoustic feedback [46]. An important class of such methods is characterized by the use of adaptive filters and, more specifically, by the use of the adaptive filters to model the unknown feedback path. Adaptive feedback cancellation (AFC) is the usual name by which these methods are identified. An illustration of a typical acoustic feedback scenario including an AFC approach is shown in fig. 5.1; the adaptive filter $\hat{F}(q,t)$ represents the estimated feedback path model which should, ideally, perfectly match the true feedback path $F_{\mathrm{t}}(q,t)$, in order to reduce the feedback artifact. $\hat{F}(q,t)$ and $F_{\mathrm{t}}(q,t)$ are assumed linear and possibly time-varying and will be further defined in section 5.2. Here, t is the discrete time index and q^{-1} is the delay operator, i. e. $q^{-k}u(t) = u(t-k)$, which allows a compact definition of the different transfer functions (TFs) and will be used throughout the paper. The nature of the problem can be seen by noticing that the microphone signal $y(t)$ is not only composed of the source signal $v(t)$, i. e. the desired signal to be amplified and sent to the loudspeaker, but also of the undesired interference $x(t)$, originating from the presence of the

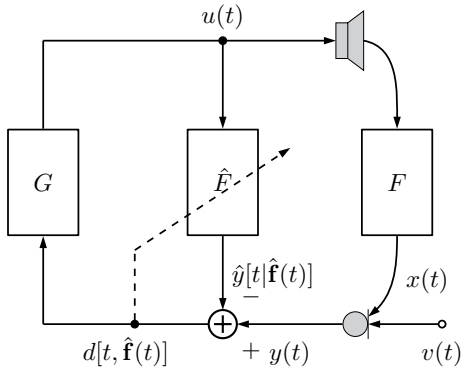


Figure 5.1: General AFC scenario.

acoustic feedback, i. e. $y(t) = x(t) + v(t)$. A similar situation also characterizes the standard acoustic echo scenario; what differentiates the acoustic feedback scenario from the acoustic echo scenario is the presence of the forward path transfer function $G(q,t)$, turning the system into a closed-loop system, and introducing a signal correlation between the loudspeaker signal $u(t)$ and the source signal $v(t)$. This correlation makes the estimation of the feedback path more problematic than in the acoustic echo scenario and, as a consequence, employing a standard adaptive filtering algorithm, e. g. the normalized least mean squares (NLMS), returns a biased estimate of $F_{\mathrm{t}}(q,t)$ [97, 83], thus limiting the cancellation properties of $\hat{F}(q,t)$. Additionally, system instabilities can be induced by the closed-loop, leading to a series of acoustic artifacts such as howling. In order to reduce these problems and obtain a reliable estimate, a procedure for decorrelating $v(t)$ and $u(t)$ should be included.

Different approaches have been proposed in the literature to reduce the signal correlation in the acoustic feedback scenario, and thus produce better estimates of the feedback path TF, such as the introduction of an external probe noise [102, 103], modifications of the forward path TF by means of nonlinear processing [104, 105], time-varying processing [104, 106] and added delays [92], two-microphones strategies [147], and, more recently, the use of a prewhitening filter used for decorrelation [148, 82, 149, 87, 88]. The latter approach relies on the use of an appropriate model for the disturbance of the identification procedure which, in the AFC context, is represented by the source signal $v(t)$.

The prewhitening filter-based AFC has been shown to be advantageous since it provides limited perceptual distortions, unlike the other aforementioned approaches [46, 121]. However, the need for a source signal model introduces a new challenge from the identification point of view, since the unknown source

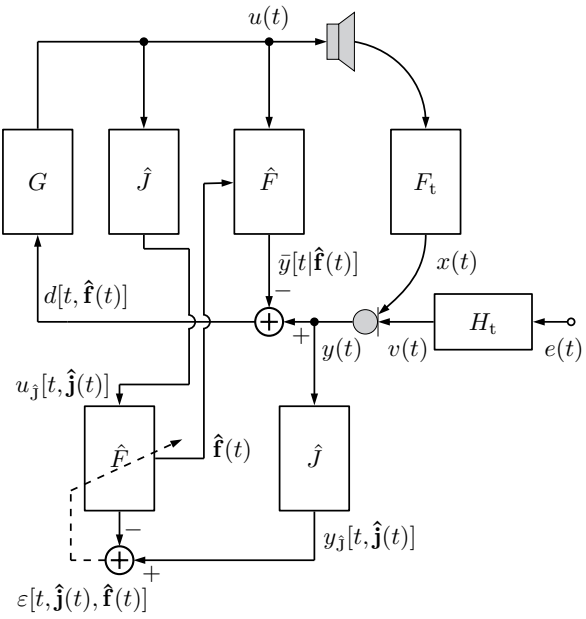


Figure 5.2: Complete AFC algorithm with PEM stage.

signal $v(t)$ is usually a nonstationary speech or audio signal. Nonstationarity implies that the source signal model for $v(t)$ must be concurrently estimated alongside the estimation of the feedback path model. Therefore, the identifiability conditions (ICs) of the system, which now includes two models to be identified, are inevitably changed [83, 109].

The application of the prediction-error method (PEM) to prewhitening filter-based AFC has been widely studied [83, 150, 110, 86, 109, 93], resulting in several different algorithms, e. g. the PEM-based adaptive filtering with row operations (PEM-AFROW), as well as interesting results regarding model identifiability. In the time-invariant case, with a true source signal generation system $H_{\mathrm{t}}(q)$ defined by an autoregressive (AR) process with a white noise excitation signal $e(t)$, see fig. 5.2, Spriet *et al.* [83] have proved that identifiability can be achieved if sufficient delay is included in the forward path or in the feedback cancellation path, as well as if a time-varying or nonlinear processing forward path TF is considered. This identifiability analysis has subsequently been extended to a wider range of source signal models [109].

The AFC has been also formulated in the frequency domain, i. e. as a frequency-domain adaptive filter (FDAF), and combined with a time-domain prewhitening

filter, i. e. the prediction-error method (PEM)-based frequency-domain adaptive filter (PEM-FDAF) [84, 110]. More recently, a PEM-based prewhitening filter has been used in combination with a frequency-domain Kalman filter (FDKF) applied to a state-space structure, leading to the so-called PEM-based frequency-domain Kalman filter (PEM-FDKF) [150], to achieve better convergence and tracking properties compared to the PEM-FDAF [84, 144]. An advantage of the PEM-FDKF is the inherent optimal choice of the step-size parameter [151] which usually needs to be fixed as design parameter of the PEM-FDAF algorithm or adaptively estimated using variable step-size algorithms [144, 87, 95].

In this paper, we provide the complete derivation of the PEM-FDKF algorithm, which was not included in [150], together with a complexity analysis and a study of the ICs for the closed-loop identification. Additionally, we propose an extension of the PEM-FDKF by means of a partitioned block (PB) frequency-domain implementation, referred to as the PEM-based partitioned-block frequency-domain Kalman filter (PEM-PBFDKF), allowing the reduction of the algorithmic delay, as needed in, e. g., hearing aid (HA) applications. The paper is organized as follows. In section 5.2, we review the PEM for direct closed-loop system identification. In section 5.3, we introduce the PEM-FDKF, providing a complete derivation of the algorithm. In section 5.4, we study the ICs, allowing to obtain a unique and unbiased model estimate for both the feedback path and the source signal generation system. In section 5.5, we present the extension of the PEM-FDKF relying on partitioned block (PB) processing, the PEM-PBFDKF. In section 5.6, we provide a computational complexity and memory requirements comparison of the proposed algorithms. In section 5.7, we illustrate the performance of the proposed algorithms in terms of convergence speed, added stability and sound quality by means of simulation results. Finally, the conclusions are drawn in section 5.8.

5.2 Prediction error method identification

The PEM is widely used in direct closed-loop system identification [97, 83]. For the case illustrated in fig. 5.2, the PEM can be used to provide a direct closed-loop identification [108, 97] of both the true feedback path $F_{\mathrm{t}}(q,t)$ and the true source signal generation system $H_{\mathrm{t}}(q,t)$. Throughout the paper, we use the following notation system: a symbol with the subscript $_{\mathrm{t}}$ refers to the true system, a regular symbol refers to the model, and a symbol with a hat refers to the model estimate; e. g., $F_{\mathrm{t}}(q,t)$ is the true feedback path, $F(q,t)$ is the feedback path model, and $\hat{F}(q,t)$ is the feedback path model estimate.

Assuming $F(q,t)$ and $H(q,t)$ to be parametric difference equation models, and defining a new model $J(q,t)$ satisfying the equation $J(q,t)H(q,t)=1$ for later

use, we introduce the parameter vectors $\boldsymbol{\theta}(t)$, $\mathbf{f}(t)$, and $\mathbf{j}(t)$:

$$\boldsymbol{\theta}(t) = [\mathbf{f}^{\mathrm{T}}(t) \quad \mathbf{j}^{\mathrm{T}}(t)]^{\mathrm{T}} \tag{5.1}$$

$$\mathbf{f}(t) = [f_0(t) \quad f_1(t) \quad \dots \quad f_{n_F-1}(t)]^{\mathrm{T}} \tag{5.2}$$

$$\mathbf{j}(t) = [1 \quad j_1(t) \quad \dots \quad j_{n_J-1}(t)]^{\mathrm{T}}, \tag{5.3}$$

where $n_\theta = n_F + n_J$. Assuming the true system is contained in the model set [108], the true system can be described using the true values of $\mathbf{f}(t)$, i. e. $\mathbf{f}_{\mathrm{t}}(t)$, as

$$F_{\mathrm{t}}(q,t) = F(q,t)\big|_{\mathbf{f}(t)=\mathbf{f}_{\mathrm{t}}(t)} \tag{5.4a}$$

$$= f_{\mathrm{t},0}(t) + f_{\mathrm{t},1}(t)q^{-1} + \ldots + f_{\mathrm{t},n_F-1}(t)q^{-n_F+1} \tag{5.4b}$$

and thus

$$y(t) = F_{\mathrm{t}}(q,t)u(t) + H_{\mathrm{t}}(q,t)e(t). \tag{5.5}$$

Similarly to (5.4), the true value of $\mathbf{j}(t)$, i. e. $\mathbf{j}_{\mathrm{t}}(t)$, can be used to write $J_{\mathrm{t}}(q,t) = J(q,t)\big|_{\mathbf{j}(t)=\mathbf{j}_{\mathrm{t}}(t)}$, with $J_{\mathrm{t}}(q,t)H_{\mathrm{t}}(q,t) = 1$.

We can now define the prediction error (PE) using the one-step ahead predictor for $y(t)$, $\bar{y}[t|\mathbf{f}(t)]$, as

$$\varepsilon[t,\boldsymbol{\theta}(t)] = J(q,t)y(t) - J(q,t)\bar{y}[t|\mathbf{f}(t)] \tag{5.6a}$$

$$= J(q,t)\left[y(t) - F(q,t)u(t)\right], \tag{5.6b}$$

and find the true values of the parameter vectors $\mathbf{f}(t)$ and $\mathbf{j}(t)$, by minimizing the variance of the PE

$$\min_{\boldsymbol{\theta}(t)} \quad \mathbb{E}\{\varepsilon^2[t,\boldsymbol{\theta}(t)]\}, \tag{5.7}$$

where $\mathbb{E}\{\cdot\}$ denotes statistical expectation and the measured $\varepsilon[t,\boldsymbol{\theta}(t)]$ is considered to be a realization of the PE, deriving from a realization of the white noise excitation $e(t)$, being the only random variable in this scenario.

The ICs, i. e. the conditions that allow to uniquely estimate $F_{\mathrm{t}}(q,t)$ and $J_{\mathrm{t}}(q,t)$, have been derived in literature [83] by converting the nonlinear PEM cost function (5.7) into a linear cost function, by means of the transformation:

$$A(q,t) = J(q,t) \tag{5.8}$$

$$B(q,t) = -J(q,t)F(q,t) \tag{5.9}$$

with $A(q,t)$ and $B(q,t)$ parameterized by

$$\boldsymbol{\xi}(t) = [\mathbf{a}^{\mathrm{T}}(t) \quad \mathbf{b}^{\mathrm{T}}(t)]^{\mathrm{T}} \tag{5.10}$$

$$\mathbf{a}(t) = [1 \quad a_1(t) \quad \dots \quad a_{n_A}(t)]^{\mathrm{T}} \tag{5.11}$$

$$\mathbf{b}(t) = [b_0(t) \quad b_1(t) \quad \dots \quad b_{n_B}(t)]^{\mathrm{T}}, \tag{5.12}$$

allowing to rewrite (5.6b) as

$$\varepsilon[t, \boldsymbol{\xi}(t)] = A(q,t)y(t) + B(q,t)u(t). \tag{5.13}$$

and (5.7) as the constrained optimization problem

$$\min_{\boldsymbol{\xi}(t)} \qquad \mathbb{E}\{\varepsilon^2[t, \boldsymbol{\xi}(t)]\} \tag{5.14a}$$

$$\text{subject to} \qquad A(q,t) \text{ is a divisor of } B(q,t), \tag{5.14b}$$

where the constraint follows from (5.8) and (5.9). The first tap parameter of $A(q,t)$ is always set to 1, i. e. $A(q,t) = 1 + q^{-1}\bar{A}(q,t)$, in order to avoid the trivial solution $A(q,t) = B(q,t) = 0$ in (5.14). The PE definition in (5.13) is used in the first part of section 5.4, where we provide the ICs for the optimization problem leading to the proposed PEM-FDKF. Unlike in the derivation in [83], here $A(q,t)$ and $B(q,t)$ are considered to be time-varying quantities. Later in section 5.4, we use again the PE definition in (5.6b) from the derivation [it will also be clear that the constraint (5.14b) may be removed] as the solution of (5.14a) alone satisfies the constraint (5.14b) and hence the solution of (5.14a) is also equal to the solution of (5.7).

In the next section, describing the algorithmic derivation of the PEM-FDKF, we also use (5.6b) to describe $\varepsilon[t, \boldsymbol{\theta}(t)]$. Following the PEM, the optimization in (5.7), w.r.t. $\mathbf{f}(t)$ and $\mathbf{j}(t)$, is carried out in an alternating fashion [93], i. e. estimating, at each iteration, in a first step the coefficients of $J(q,t)$ with fixed estimates for $F(q,t)$ and in a second step the coefficients of $F(q,t)$ with fixed estimates for $J(q,t)$.

5.3 PEM-Based Frequency Domain Kalman Filter

The PEM-FDKF algorithm [150] is an extension of the algorithm proposed by Enzner and Vary [151] for acoustic echo cancellation (AEC). It relies on a dynamical model for the feedback path and a model for the recorded microphone signal to define a state-space representation to which the Kalman filter procedure

Table 5.1: Definition of the time-invariant constraints used in the paper, as defined in Benesty *et al.* [146].

Constraint rectangular matrix	Constraint square matrix
$\mathbf{W}^{01}_{R\times M} = \begin{bmatrix}\mathbf{0}_{R\times R} & \mathbf{I}_R\end{bmatrix}$	$\mathbf{W}^{01}_{M\times M} = \begin{bmatrix}\mathbf{0}_{R\times R} & \mathbf{0}_{R\times R}\\ \mathbf{0}_{R\times R} & \mathbf{I}_R\end{bmatrix}$
$\mathbf{W}^{10}_{M\times R} = \begin{bmatrix}\mathbf{I}_R\\ \mathbf{0}_{R\times R}\end{bmatrix}$	$\mathbf{W}^{10}_{M\times M} = \begin{bmatrix}\mathbf{I}_R & \mathbf{0}_{R\times R}\\ \mathbf{0}_{R\times R} & \mathbf{0}_{R\times R}\end{bmatrix}$

can be applied. The simple frequency-domain dynamical model chosen for the feedback path employs a first-order Markov model as an abstraction of the true feedback path dynamics. Similar models have also been proposed in the time-domain [98, 152, 153]; however, the use of a block-wise procedure has an impact on the calculation of the model time constant. The main change introduced to the algorithm by Enzner and Vary consists of the decorrelation stage, by means of a PEM-based prewhitening filter already seen in the PEM-FDAF [143, 144, 154]. In this way, the frequency-domain Kalman filter (FDKF) framework can be successfully applied to AFC, leading to the PEM-FDKF.

The two main advantages of a frequency-domain approach are the lower computational complexity and the good decorrelation properties of the discrete Fourier transform (DFT) [113, 155, 110]. In HA applications, where short filters are usually employed, the computational complexity advantage is smaller than in PA applications, but it can be still relevant, as we will show in section 5.6, especially if the DFT calculations are shared by other processing stages of the HA. In addition to these two advantages, the formulation of the problem based on a state-space representation allows the optimal estimation of the FDAF stepsize as part of the Kalman filter procedure [151, 156], leading to a significantly improved convergence [150].

It should be emphasized that this approach employs the implicit assumption that $F_{\mathrm{t}}(q,t)$ and $H_{\mathrm{t}}(q,t)$ are slowly time-varying [with $F_{\mathrm{t}}(q,t)$ varying more slowly than $H_{\mathrm{t}}(q,t)$], since these are modeled by $F(q,\kappa)$ and $H(q,\kappa)$, where $\kappa \in \mathbb{Z}$ is the time frame index, which hence can only vary at the frame rate [151, 157]. Therefore, we will assume time invariance of $F_{\mathrm{t}}(q,t)$ and $H_{\mathrm{t}}(q,t)$ over each frame, i. e. effectively over each frame shift of R samples, and use the notation $F_{\mathrm{t}}(q,\kappa)$ and $H_{\mathrm{t}}(q,\kappa)$. Similarly, $J_{\mathrm{t}}(q,\kappa)$ is defined such that $J_{\mathrm{t}}(q,\kappa)H_{\mathrm{t}}(q,\kappa) = 1$.

The discrete frequency index l will be used, together with the frame index κ, to describe the time-frequency components of the different variables. The first

Table 5.2: Definition of linearization matrices used in the paper, as defined in Benesty *et al.* [146].

Linearization rectangular matrix	Linearization square matrix
$\mathbf{G}^{01}_{R\times M} = \mathbf{F}_R\,\mathbf{W}^{01}_{R\times M}\,\mathbf{F}_M^{-1}$	$\mathbf{G}^{01}_{M\times M} = (\mathbf{G}^{01}_{R\times M})^{\mathrm{H}}\mathbf{G}^{01}_{R\times M}$ $= \mathbf{F}_M\,\mathbf{W}^{01}_{M\times M}\,\mathbf{F}_M^{-1}$
$\mathbf{G}^{10}_{M\times R} = \mathbf{F}_M\,\mathbf{W}^{10}_{M\times R}\,\mathbf{F}_R^{-1}$	$\mathbf{G}^{10}_{M\times M} = \mathbf{G}^{10}_{M\times R}\,(\mathbf{G}^{10}_{M\times R})^{\mathrm{H}}$ $= \mathbf{F}_M\,\mathbf{W}^{10}_{M\times M}\,\mathbf{F}_M^{-1}$

introduced variable is the M-samples loudspeaker signal for frame κ,

$$\mathbf{u}(\kappa) = [u(\kappa R - M + 1) \quad \dots \quad u(\kappa R)]^{\mathrm{T}}, \tag{5.15}$$

where R denotes the frame shift.

Assuming the true value $J_{\mathrm{t}}(q,\kappa)$ to be available, the prefiltered version of the loudspeaker signal for frame κ is

$$\begin{aligned}\mathbf{u}_{\mathrm{J_t}}(\kappa) &= [J_{\mathrm{t}}(q,\kappa)u(\kappa R - M + 1) \ \dots \ J_{\mathrm{t}}(q,\kappa)u(\kappa R)]^{\mathrm{T}} \\ &= [u_{\mathrm{J_t}}(\kappa R - M + 1) \quad \dots \quad u_{\mathrm{J_t}}(\kappa R)]^{\mathrm{T}}.\end{aligned} \tag{5.16}$$

The frequency-domain version of the prefiltered loudspeaker signal is then given in matrix form as

$$\mathbf{U}_{\mathrm{J_t}}(\kappa) = \mathrm{diag}\left\{\mathbf{F}_M\,\mathbf{u}_{\mathrm{J_t}}(\kappa)\right\}, \tag{5.17}$$

where $\mathbf{F}_M$ is the unitary DFT matrix of size $M \times M$, i. e. $\mathbf{F}_M^{-1} = \mathbf{F}_M^{\mathrm{H}}$, and the diag$\{\cdot\}$ operator either maps an $M \times 1$ vector to the diagonal of an $M \times M$ diagonal matrix, or maps an $M \times M$ matrix to the $M \times 1$ vector given by its diagonal. The dimension parameters R and M should be chosen properly, taking into consideration the length of the true feedback path or an estimate thereof. A common choice is $R = n_{\hat{F}}$ (assuming $n_{\hat{F}} = n_F$) and $M = 2R$ [113, 144]; if the algorithmic delay, equal to $2R - 1$, is not acceptable, a PB solution (also known as multidelay filter) [115, 155, 84] can be chosen, see section 5.5.

With $R = n_{\hat{F}}$, the frequency-domain version of the true feedback path parameter vector $\mathbf{f}_{\mathrm{t}}(\kappa)$ is

$$\mathbf{F}_{\mathrm{t}}(\kappa) = \mathbf{G}^{10}_{M\times R}\,\mathbf{F}_R\,\mathbf{f}_{\mathrm{t}}(\kappa) \tag{5.18a}$$

$$= \mathbf{F}_M\,\mathbf{W}^{10}_{M\times R}\,\mathbf{f}_{\mathrm{t}}(\kappa). \tag{5.18b}$$

The rectangular matrices $\mathbf{G}^{10}_{M\times R}$ and $\mathbf{W}^{10}_{M\times R}$ are used to obtain the $M \times 1$ frequency-domain version of the $R \times 1$ vector $\mathbf{f}_\mathrm{t}(\kappa)$. The smaller of the two matrix dimensions R always indicates the dimension of the identity matrix $\mathbf{I}_R$ appearing in the matrix definitions. The simplified definitions of these matrices, considering the standard case $M = 2R$, are shown in tables 5.1 and 5.2 along with the other matrices needed to compactly describe the algorithm in the frequency-domain as defined by Benesty *et al.* [146]. Despite the two classes of definitions for frequency-domain quantities, e. g. using (5.18a) or (5.18b), we will only use the class of definitions including the linearization matrices $\mathbf{G}^{\cdot\cdot}_{\cdot\times\cdot}$ in the rest of the paper.

Finally, we introduce the R-samples prefiltered microphone signal and the source excitation signal for frame κ, i. e.

$$\begin{aligned}\mathbf{y}_{\mathrm{J_t}}(\kappa) &= [J_\mathrm{t}(q,\kappa)y(\kappa R - R + 1) \;\ldots\; J_\mathrm{t}(q,\kappa)y(\kappa R)]^\mathrm{T} \\ &= [y_{\mathrm{J_t}}(\kappa R - R + 1) \quad \ldots \quad y_{\mathrm{J_t}}(\kappa R)]^\mathrm{T} \end{aligned} \tag{5.19}$$

$$\mathbf{e}(\kappa) = [e(\kappa R - R + 1) \quad \ldots \quad e(\kappa R)]^\mathrm{T}, \tag{5.20}$$

and their frequency-domain versions:

$$\mathbf{Y}_{\mathrm{J_t}}(\kappa) = (\mathbf{G}^{01}_{R\times M})^\mathrm{H}\mathbf{F}_R\,\mathbf{y}_{\mathrm{J_t}}(\kappa) \tag{5.21}$$

$$\mathbf{E}(\kappa) = (\mathbf{G}^{01}_{R\times M})^\mathrm{H}\mathbf{F}_R\,\mathbf{e}(\kappa). \tag{5.22}$$

The quantities introduced so far can be combined into a state-space representation using the frequency-domain Markov model for the feedback path $\mathbf{F}_\mathrm{t}(\kappa)$ as a state equation and the linear model for the frequency-domain prefiltered microphone signal $\mathbf{Y}_{\mathrm{J_t}}(\kappa)$ as a measurement equation:

$$\mathbf{F}_\mathrm{t}(\kappa+1) = \alpha_\mathrm{t}\mathbf{F}_\mathrm{t}(\kappa) + \mathbf{N}_\mathrm{t}(\kappa) \tag{5.23a}$$

$$\mathbf{Y}_{\mathrm{J_t}}(\kappa) = \mathbf{C}_{\mathrm{J_t}}(\kappa)\mathbf{F}_\mathrm{t}(\kappa) + \mathbf{E}(\kappa) \tag{5.23b}$$

where $\mathbf{C}_{\mathrm{J_t}}(\kappa) = \mathbf{G}^{01}_{M\times M}\,\mathbf{U}_{\mathrm{J_t}}(\kappa)$ includes the linear transformation to be applied to $\mathbf{U}_{\mathrm{J_t}}(\kappa)$ to linearize the circular convolution between $\mathbf{u}_{\mathrm{J_t}}(\kappa)$ and $\mathbf{f}_\mathrm{t}(\kappa)$, cf. table 5.2, $\mathbf{N}_\mathrm{t}(\kappa)$ is the process noise describing the unpredictability of the feedback path dynamics and α_t is the transition factor accounting for the time-variability of the feedback path [151]. The total number of linear convolution samples resulting from the circular convolution, given the chosen signal dimension parameters, is $M - R + 1 = R + 1$ [113]. However, we retain only R samples to match the frame shift and simplify the notation, as commonly done in the literature [113, 146, 151, 158].

Assuming $J_{\mathrm{t}}(q,\kappa)$ is indeed available, the use of prefiltered variables in (5.23b) guarantees the decorrelation between the prefiltered loudspeaker signal $\mathbf{u}_{\mathrm{J_t}}(\kappa)$ and the source excitation signal $\mathbf{e}(\kappa)$ in the measurement equation, thus achieving the necessary requirements to employ a Kalman filter for the estimation of $\mathbf{F}_{\mathrm{t}}(\kappa)$. The linear minimum mean-square error (MMSE) estimate of the state vector $\mathbf{F}_{\mathrm{t}}(\kappa)$ corresponds to the solution of a Bayesian optimization problem [159, ch. 13], and is given by the well-known set of equations referred to as the (block) Kalman filter:

$$\mathbf{K}(\kappa) = \mathbf{P}(\kappa)\mathbf{C}_{\mathrm{J_t}}^{\mathrm{H}}(\kappa)\left[\mathbf{C}_{\mathrm{J_t}}(\kappa)\mathbf{P}(\kappa)\mathbf{C}_{\mathrm{J_t}}^{\mathrm{H}}(\kappa) + \mathbf{\Psi}_{\mathrm{EE}}(\kappa)\right]^{-1} \tag{5.24a}$$

$$\hat{\mathbf{F}}^{+}(\kappa) = \hat{\mathbf{F}}(\kappa) + \mathbf{K}(\kappa)[\mathbf{Y}_{\mathrm{J_t}}(\kappa) - \mathbf{C}_{\mathrm{J_t}}(\kappa)\hat{\mathbf{F}}(\kappa)] \tag{5.24b}$$

$$\mathbf{P}^{+}(\kappa) = [\mathbf{I}_M - \mathbf{K}(\kappa)\mathbf{C}_{\mathrm{J_t}}(\kappa)]\,\mathbf{P}(\kappa) \tag{5.24c}$$

$$\hat{\mathbf{F}}(\kappa+1) = \alpha_{\mathrm{t}} \cdot \hat{\mathbf{F}}^{+}(\kappa) \tag{5.24d}$$

$$\mathbf{P}(\kappa+1) = \alpha_{\mathrm{t}}^2 \cdot \mathbf{P}^{+}(\kappa) + \mathbf{\Psi}_{\mathrm{N_tN_t}}(\kappa), \tag{5.24e}$$

where $\mathbf{K}(\kappa)$ is the frequency-domain Kalman gain, the superscript $^{+}$ indicates a posteriori estimates and, finally, $\mathbf{\Psi}_{\mathrm{EE}}(\kappa)$ and $\mathbf{\Psi}_{\mathrm{N_tN_t}}(\kappa)$ are the covariance matrices of $\mathbf{E}(\kappa)$ and $\mathbf{N}_{\mathrm{t}}(\kappa)$ [157, 151], respectively, assumed to be known.

For implementation purposes, we drop some of the assumptions initially made for the state space model, similarly to what is done in the PEM-FDAF in order to carry out the optimization in an alternating fashion [84]:

1. The prefiltering operation is performed by means of the estimated $\hat{J}(q,\kappa)$, instead of the true $J_{\mathrm{t}}(q,\kappa)$; therefore, the prefiltered variables will be, from now on, indicated using the subscript $_{\hat{\mathrm{J}}}$, instead of $_{\mathrm{J_t}}$. Following a common assumption found in literature [70, 84, 99], the system $H_{\mathrm{t}}(q,\kappa)$ generating the source signal $v(t)$ is assumed to be time-varying, monic, inversely stable, and AR. The $J_{\mathrm{t}}(q,\kappa)$ is then estimated as the linear prediction filter for the time-domain error signal frame, i. e. $\mathbf{d}[\kappa|\hat{\mathbf{f}}(\kappa)] = \left[d[\kappa R - R + 1|\hat{\mathbf{f}}(\kappa)] \;\ldots\; d[\kappa R|\hat{\mathbf{f}}(\kappa)]\right]^{\mathrm{T}}$ at the current and the previous frame, i. e. $[\mathbf{d}^{\mathrm{T}}[\kappa|\hat{\mathbf{f}}(\kappa)]\;\mathbf{d}^{\mathrm{T}}[\kappa-1|\hat{\mathbf{f}}(\kappa-1)]]^{\mathrm{T}}$, using the Levinson-Durbin algorithm [90, pp. 254-264] and represented by $\hat{\mathbf{j}}(\kappa)$.
2. The use of estimated prewhitening filter parameters motivates the replacement of $[\mathbf{Y}_{\mathrm{J_t}}(\kappa) - \mathbf{C}_{\mathrm{J_t}}(\kappa)\hat{\mathbf{F}}(\kappa)]$ in (5.24b) by the frequency-domain

PE frame $\boldsymbol{\mathcal{E}}[\kappa, \hat{\boldsymbol{\Theta}}(\kappa)]$, related to the time-domain PE frame $\boldsymbol{\varepsilon}[\kappa, \hat{\boldsymbol{\theta}}(\kappa)]$ via

$$\boldsymbol{\varepsilon}[\kappa, \hat{\boldsymbol{\theta}}(\kappa)] = \Big[\varepsilon[\kappa R - R + 1, \hat{\boldsymbol{\theta}}(\kappa)] \ \ldots \ \varepsilon[\kappa R, \hat{\boldsymbol{\theta}}(\kappa)]\Big]^{\mathrm{T}} \tag{5.25}$$

$$\boldsymbol{\mathcal{E}}[\kappa, \hat{\boldsymbol{\Theta}}(\kappa)] = (\mathbf{G}^{01}_{R\times M})^{\mathrm{H}} \mathbf{F}_R \, \boldsymbol{\varepsilon}[\kappa, \hat{\boldsymbol{\theta}}(\kappa)] \tag{5.26}$$

with $\hat{\boldsymbol{\Theta}}(\kappa)$ containing the frequency-domain versions of $\hat{\mathbf{f}}(\kappa)$ and $\hat{\mathbf{j}}(\kappa)$ [cf. (5.46) to (5.50)]. Furthermore, we add the linearization constraint $\mathbf{G}^{10}_{M\times M}$ (cf. table 5.2) in (5.24b) [cf. (5.31b)], similarly to what is done in the FDAFs, given the improved sound quality provided by a constrained FDAF version [113]. Naturally, such a constraint causes a computational complexity increase.

Additionally, we use the approximations introduced by Enzner and Vary [151], in order to address both the problem of high computational complexity and the possible ill-posedness of the solution, via diagonal operations, as follows:

1. The linearization square matrix $\mathbf{G}^{01}_{M\times M}$ in the definition of $\mathbf{C}_{\hat{\mathrm{j}}}(\kappa)$ is approximated by a diagonal matrix, allowing to write $\mathbf{G}^{01}_{M\times M} \approx (R/M)\mathbf{I}_M$ and, less intuitively, $\mathbf{G}^{01}_{M\times M}\, \boldsymbol{\Delta} (\mathbf{G}^{01}_{M\times M})^{\mathrm{H}} \approx (R/M)\boldsymbol{\Delta}$, if $\boldsymbol{\Delta}$ is a diagonal matrix [146, ch. 8].
2. The covariance matrices $\boldsymbol{\Psi}_{\mathrm{EE}}(\kappa)$ and $\boldsymbol{\Psi}_{\mathrm{N_t N_t}}(\kappa)$ are replaced by the estimates $\boldsymbol{\Psi}_{\hat{\mathrm{E}}\hat{\mathrm{E}}}(\kappa)$ and $\boldsymbol{\Psi}_{\hat{\mathrm{N}}\hat{\mathrm{N}}}(\kappa)$, respectively, which are assumed to be diagonal [151]. These are related to the corresponding time-varying power spectral densities $\boldsymbol{\Phi}_{\hat{\mathrm{E}}\hat{\mathrm{E}}}(\kappa)$ and $\boldsymbol{\Phi}_{\hat{\mathrm{N}}\hat{\mathrm{N}}}(\kappa)$, adaptively estimated using the procedures described in [157, 160, 158], assuming that the feedback path is slowly time varying.
3. Given the assumed low correlation between different frequency components of the estimation error [151], the matrix $\mathbf{P}(\kappa)$ is a nearly diagonal matrix; the diagonalilty is enforced by initalizing $\mathbf{P}(0) \propto \mathbf{I}_M$.

The last three approximations are used to write the following simplified expressions:

$$\mathbf{C}_{\hat{\mathrm{j}}}(\kappa) \approx \frac{R}{M} \cdot \mathbf{U}_{\hat{\mathrm{j}}}(\kappa) \tag{5.27}$$

$$\mathbf{C}_{\hat{\mathrm{j}}}(\kappa)\mathbf{P}(\kappa)\mathbf{C}^{\mathrm{H}}_{\hat{\mathrm{j}}}(\kappa) \approx \frac{R}{M} \cdot \mathbf{U}_{\hat{\mathrm{j}}}(\kappa)\mathbf{P}(\kappa)\mathbf{U}^{\mathrm{H}}_{\hat{\mathrm{j}}}(\kappa) \tag{5.28}$$

$$\boldsymbol{\Psi}_{\hat{\mathrm{E}}\hat{\mathrm{E}}}(\kappa) \approx R \cdot \mathrm{diag}\{\boldsymbol{\Phi}_{\hat{\mathrm{E}}\hat{\mathrm{E}}}(\kappa)\} \tag{5.29}$$

$$\boldsymbol{\Psi}_{\hat{\mathrm{N}}\hat{\mathrm{N}}}(\kappa) \approx M \cdot \mathrm{diag}\{\boldsymbol{\Phi}_{\hat{\mathrm{N}}\hat{\mathrm{N}}}(\kappa)\}. \tag{5.30}$$

With the approximations discussed so far, the complete set of equations describing the PEM-FDKF update is as follows:

$$\mathbf{K}(\kappa) = \mathbf{P}(\kappa)\mathbf{U}_{\hat{\mathrm{j}}}^{\mathrm{H}}(\kappa)\left[\mathbf{U}_{\hat{\mathrm{j}}}(\kappa)\mathbf{P}(\kappa)\mathbf{U}_{\hat{\mathrm{j}}}^{\mathrm{H}}(\kappa) + M \cdot \mathrm{diag}\{\mathbf{\Phi}_{\hat{\mathrm{E}}\hat{\mathrm{E}}}(\kappa)\}\right]^{-1} \tag{5.31a}$$

$$\hat{\mathbf{F}}^{+}(\kappa) = \hat{\mathbf{F}}(\kappa) + \mathbf{G}_{M\times M}^{10}\,\mathbf{K}(\kappa)\boldsymbol{\mathcal{E}}[\kappa, \hat{\mathbf{\Theta}}(\kappa)] \tag{5.31b}$$

$$\mathbf{P}^{+}(\kappa) = \left[\mathbf{I}_M - \frac{R}{M}\mathbf{K}(\kappa)\mathbf{U}_{\hat{\mathrm{j}}}(\kappa)\right]\mathbf{P}(\kappa) \tag{5.31c}$$

$$\hat{\mathbf{F}}(\kappa+1) = \alpha \cdot \hat{\mathbf{F}}^{+}(\kappa) \tag{5.31d}$$

$$\mathbf{P}(\kappa+1) = \alpha^2 \cdot \mathbf{P}^{+}(\kappa) + M \cdot \mathrm{diag}\{\mathbf{\Phi}_{\hat{\mathrm{N}}\hat{\mathrm{N}}}(\kappa)\}. \tag{5.31e}$$

5.4 PEM-FDKF Identifiability conditions

The ICs for the optimization problem solved by the PEM-FDKF will be derived in three steps, as follows.

The first step, providing ICs for (5.7) and (5.14), is similar to the derivation in [83], but now introducing time variability in the signal models. The following expressions for $y(t)$ and $u(t)$ hold (cf. fig. 5.2):

$$y(t) = F_{\mathrm{t}}(q,t)u(t) + v(t) \tag{5.32}$$

$$u(t) = G(q)\left[y(t) - \hat{F}(q,t)u(t)\right], \tag{5.33}$$

where $G(q) = q^{-d_G}\bar{G}(q)$ and $d_G \geq 1$, i. e. the forward path has to have at least a one-sample delay to avoid a delay-less loop. Using (5.32) and (5.33), the PE in (5.13) can be rewritten as

$$\begin{aligned}\varepsilon[t, \boldsymbol{\xi}(t)] &= A(q,t)v(t) + [A(q,t)F_{\mathrm{t}}(q,t) + B(q,t)] \times u(t) \\ &= A(q,t)v(t) + [A(q,t)F_{\mathrm{t}}(q,t) + B(q,t)] \\ &\quad \times q^{-d_G}\bar{G}(q)\left[y(t) - \hat{F}(q,t)u(t)\right].\end{aligned} \tag{5.34}$$

The PE $\varepsilon[t, \boldsymbol{\xi}(t)]$ can be expressed as a function of $v(t)$ only, by using (5.32) and (5.33) repeatedly, in (5.34), i. e.

$$\varepsilon[t, \boldsymbol{\xi}(t)] = A(q,t)v(t) + Z(q,t)\bar{G}(q)v(t-d_G) + Z(q,t)\sum_{\ell=2}^{\infty} \bar{G}^{\ell}(q) \left[\prod_{i=1}^{\ell-1} F_{\mathrm{r}}(q, t-id_G)\right] \times v(t-\ell d_G) \tag{5.35}$$

where

$$F_{\mathrm{r}}(q,t) = F_{\mathrm{t}}(q,t) - \hat{F}(q,t) \tag{5.36}$$

$$Z(q,t) = A(q,t)F_{\mathrm{t}}(q,t) + B(q,t). \tag{5.37}$$

As is done for the time-invariant case in [83], we consider the sufficient order condition for $A(q,t)$ and $B(q,t)$, i. e. $n_A \geq n_J$ and $n_B \geq n_J + n_F - 1$, and the causality of $\bar{G}(q)$, $F_{\mathrm{t}}(q,t)$, $J_{\mathrm{t}}(q,t)$ and $\hat{F}(q,t)$ and study the conditions under which the minimization in (5.14a) leads to the unique solution

$$A(q,t) = J_{\mathrm{t}}(q,t) \tag{5.38}$$

$$B(q,t) = -J_{\mathrm{t}}(q,t)F_{\mathrm{t}}(q,t). \tag{5.39}$$

This solution satisfies (5.14b), hence making this constraint indeed redundant.

The unique desired solution can be derived if at least one of the following conditions, similar to the conditions found in the time-invariant case [83], is fulfilled:

C1 The forward path delay d_G satisfies $d_G \geq n_A$;

C2 The cancellation path delay d_F, where $B(q,t) = q^{-d_F}\bar{B}(q,t)$ and hence $F_{\mathrm{t}}(q,t) = q^{-d_F}\bar{F}_{\mathrm{t}}(q,t)$, satisfies $d_G + d_F \geq n_A$;

C3 The TF $\bar{G}(q)$ is nonlinear.

The proofs resemble those in [83], with the difference that the assumed time variability of the signal models does not allow to compactly rewrite (5.35). Specifically, the first condition, C1, turns the minimization in (5.14) into a linear prediction of $v(t)$, i. e. $A(q,t) = J_{\mathrm{t}}(q,t)$, given that $a_0(t) = 1$. Additionally, since $n_A \geq n_J$, the equation $Z(q,t) = 0$ must hold, leading to $B(q,t) = -J_{\mathrm{t}}(q,t)F_{\mathrm{t}}(q,t)$. The second condition, C2, can be proved similarly, by including $B(q,t) = q^{-d_F}\bar{B}(q,t)$, and hence $F_{\mathrm{t}}(q,t) = q^{-d_F}\bar{F}_{\mathrm{t}}(q,t)$, in (5.35),

i. e.

$$\varepsilon[t,\boldsymbol{\xi}(t)] = A(q,t)v(t) + q^{-d_G-d_F}\left[A(q,t)\bar{F}_{\mathrm{t}}(q,t) + \bar{B}(q,t)\right] \times \bar{G}(q)\left[y(t) - \hat{F}(q,t)u(t)\right]. \tag{5.40}$$

If $d_G + d_F \geq n_A$, the unique solution of (5.40) is given by (5.38) and (5.39). The last condition, C3, can be understood considering that a nonlinear $\bar{G}(q)$ introduces additional decorrelation between the first and the other terms in (5.35), thus decoupling these terms in the minimization of (5.7). Such decoupled minimization yields, again, the values of $A(q,t)$ and $B(q,t)$ in (5.38) and (5.39). Following these results, which hold for any time-varying behavior of $J_{\mathrm{t}}(q,t)$ and $F_{\mathrm{t}}(q,t)$ fulfilling the initial assumptions as well as the ICs, we can go back to the simplified notation of the unconstrained optimization problem in (5.7), parameterized in $\boldsymbol{\theta}(t)$. Under the same ICs, the minimization in (5.7) then leads to the solution $J(q,t) = J_{\mathrm{t}}(q,t)$ and $F(q,t) = F_{\mathrm{t}}(q,t)$ [the equivalent of (5.38) and (5.39)].

The second step describes the transition from the current optimization problem (5.7) to a new optimization problem including the specific model (5.23), for which then the Kalman filter (alternating with the Levinson-Durbin algorithm) is seen to provide a suitable algorithm, and derive the ICs for this new optimization problem. We start by expressing (5.7) as a length-R frame-based expression, through the frame index κ. Assuming time invariance of $F_{\mathrm{t}}(q,t)$ and $H_{\mathrm{t}}(q,t)$ over each frame, the sample-based estimation of the parameter vector $\boldsymbol{\theta}(t)$ is replaced by the frame-based estimation of the parameter vector $\boldsymbol{\theta}(\kappa)$, found by solving at each κ

$$\min_{\boldsymbol{\theta}(\kappa)} \quad \mathbb{E}\{\|\boldsymbol{\varepsilon}[\kappa,\boldsymbol{\theta}(\kappa)]\|^2_{\boldsymbol{\zeta}^{-1}(\kappa)}\}, \tag{5.41}$$

where $\|\mathbf{z}\|_{\mathbf{W}} = \sqrt{\mathbf{z}^{\mathrm{H}}\mathbf{W}\mathbf{z}} = \left\|\mathbf{W}^{1/2}\mathbf{z}\right\|_2$ is the weighted norm of $\mathbf{z}$ induced by the positive definite matrix $\mathbf{W}$, and $\boldsymbol{\zeta}^{-1}(\kappa) = \zeta^{-1}(\kappa)\mathbf{I}_R$ will be used to compensate for power variations in the excitation signal $\mathbf{e}(\kappa)$. $\boldsymbol{\varepsilon}[\kappa,\boldsymbol{\theta}(\kappa)]$ is the length-R PE frame obtained from (5.6b) assuming the parameter vector $\boldsymbol{\theta}(\kappa)$ is constant in $t = \kappa R - R + 1, \ldots, \kappa R$. Under the same time-invariance assumption, it has been pointed out in [113] that minimizing the frame-based cost function vs. the sample-based cost function leads to the same mean-square error performance; therefore, the same ICs hold for the (5.41) and (5.7).

We now assume a frame-based state-space model for the true system using a simple Markov model to describe the state (i. e. the true feedback path) dynamics [82], and the linear model for the time-domain prefiltered microphone

signal frame $\mathbf{y}_{\mathrm{J_t}}(t)$ as the measurement equation:

$$\mathbf{f}_{\mathrm{t}}(\kappa+1) = \alpha_{\mathrm{t}}\mathbf{f}_{\mathrm{t}}(\kappa) + \mathbf{n}_{\mathrm{t}}(\kappa) \tag{5.42a}$$

$$\mathbf{y}_{\mathrm{J_t}}(\kappa) = F_{\mathrm{t}}(q,\kappa)\mathbf{u}_{\mathrm{J_t}}(\kappa) + \mathbf{e}(\kappa), \tag{5.42b}$$

cf. (5.16), (5.19) and (5.20), and we describe the random variables of (5.42) as $\mathbf{n}_{\mathrm{t}}(\kappa) \sim \mathcal{N}(0, \boldsymbol{\Lambda}_{\mathrm{t}}(\kappa))$, $\mathbf{e}(\kappa) \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\mathrm{t}}(\kappa))$ and $\mathbf{f}_{\mathrm{t}}(0) \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{f}_{\mathrm{t}}(0)}, \boldsymbol{\Pi}_{\mathrm{t}}(0))$, where $\mathcal{N}(\cdot,\cdot)$ indicates a normal distribution with specified mean and covariance.

The Kalman filter corresponding to the state-space model (5.42) effectively solves the optimization problem [161, 162, 163]

$$\min_{\{\mathbf{f}(n),\,\mathbf{n}(n)\}_{n=0}^{\kappa}} \quad \frac{1}{2}\left\|\mathbf{f}(0) - \boldsymbol{\mu}_{\mathbf{f}_{\mathrm{t}}(0)}\right\|^2_{\boldsymbol{\Pi}_{\mathrm{t}}^{-1}(0)} + \frac{1}{2}\sum_{n=0}^{\kappa-1}\|\mathbf{n}(n)\|^2_{\boldsymbol{\Lambda}_{\mathrm{t}}^{-1}(n)}$$

$$+ \frac{1}{2}\sum_{n=0}^{\kappa}\|\boldsymbol{\varepsilon}[n,\boldsymbol{\theta}(n)]\|^2_{\boldsymbol{\Sigma}_{\mathrm{t}}^{-1}(n)} \tag{5.43a}$$

$$\text{subject to} \qquad \mathbf{f}(n+1) = \alpha_{\mathrm{t}}\mathbf{f}(n) + \mathbf{n}(n). \tag{5.43b}$$

This formulation combines the different optimization problems in (5.41) for successive frames $n = 0, \ldots, \kappa$ into a single optimization framework, with the constraint (5.43b) defining the time evolution of $\mathbf{f}(n)$, and three terms characterizing the cost function:

1. A regularization term depending on $\mathbf{f}(0)$ and on the matrix $\boldsymbol{\Pi}_{\mathrm{t}}(0) = \mathbb{E}\{\left[\mathbf{f}(0) - \boldsymbol{\mu}_{\mathbf{f}_{\mathrm{t}}(0)}\right]\left[\mathbf{f}(0) - \boldsymbol{\mu}_{\mathbf{f}_{\mathrm{t}}(0)}\right]^{\mathrm{T}}\}$, with $\boldsymbol{\mu}_{\mathbf{f}_{\mathrm{t}}(0)}$ an initial guess of the initial state $\mathbf{f}(0)$;
2. A term depending on the unknown state noise process $\mathbf{n}(n)$, with $\mathbf{n}(n)$ being a new variable to optimize, and on $\boldsymbol{\Lambda}_{\mathrm{t}}(n) = \mathbb{E}\{\mathbf{n}(n)\mathbf{n}(n)^{\mathrm{T}}\}$;
3. A term depending on the PE $\boldsymbol{\varepsilon}[n,\boldsymbol{\theta}(n)]$, similar to the term in (5.41), where the expectation on the single frame is replaced by a summation over the different frames, and $\boldsymbol{\zeta}^{-1}(\kappa)$ in (5.41) is replaced by the matrix $\boldsymbol{\Sigma}_{\mathrm{t}}(n)$ [164, 161].

With (5.43) effectively $\mathbf{f}(n)$ is estimated under the assumption that $\mathbf{j}(n)$ is known [as $\mathbf{j}(n)$ is included in $\boldsymbol{\theta}(n)$ in the third term of (5.43a)]. If $\mathbf{j}(n) = \mathbf{j}_{\mathrm{t}}(n)$, then the Kalman filter (subject to technical full-rank conditions) provides the unique MMSE estimate of $\mathbf{f}_{\mathrm{t}}(\kappa)$, which itself (subject to the above ICs) is included in the unique minimizer of the third term of (5.43a) with the expectation reintroduced.

When the Kalman Filter operations [to estimate $\mathbf{f}(n)$] are alternated with the Levinson-Durbin algorithm [to estimate $\mathbf{j}(n)$], as in section 5.3, the optimization problem that is effectively solved is

$$\min_{\{\boldsymbol{\theta}(n),\,\mathbf{n}(n)\}_{n=0}^{\kappa}} \quad \frac{1}{2}\left\|\mathbf{f}(0)-\boldsymbol{\mu}_{\mathbf{f}_{\mathrm{t}}(0)}\right\|^2_{\boldsymbol{\Pi}_{\mathrm{t}}^{-1}(0)}+\frac{1}{2}\sum_{n=0}^{\kappa-1}\|\mathbf{n}(n)\|^2_{\boldsymbol{\Lambda}_{\mathrm{t}}^{-1}(n)}$$

$$+\frac{1}{2}\sum_{n=0}^{\kappa}\|\boldsymbol{\varepsilon}[n,\boldsymbol{\theta}(n)]\|^2_{\boldsymbol{\Sigma}_{\mathrm{t}}^{-1}(n)} \tag{5.44a}$$

$$\text{subject to} \quad \mathbf{f}(n+1)=\alpha_{\mathrm{t}}\mathbf{f}(n)+\mathbf{n}(n). \tag{5.44b}$$

The alternation minimization then provides a (possibly suboptimal) estimate of $\boldsymbol{\theta}_{\mathrm{t}}(\kappa)=[\mathbf{f}_{\mathrm{t}}^{\mathrm{T}}(\kappa)\;\mathbf{j}_{\mathrm{t}}^{\mathrm{T}}(\kappa)]^{\mathrm{T}}$, which itself (subject to the above ICs) is included in the unique minimizer of the third term of (5.44a) with the expectation reintroduced. If the Kalman filter is applied with $\mathbf{j}(n)=\mathbf{j}_{\mathrm{t}}(n)$, then (subject to technical full-rank conditions) it provides the unique MMSE estimate of $\mathbf{f}_{\mathrm{t}}(\kappa)$. Similarly, if the Levinson-Durbin algorithm is applied with $\mathbf{f}(n)=\mathbf{f}_{\mathrm{t}}(n)$ it provides the unique MMSE estimate of $\mathbf{j}_{\mathrm{t}}(\kappa)$.

The third step is that of formulating the optimization problem in the frequency domain, i. e.

$$\min_{\{\boldsymbol{\Theta}(n),\mathbf{N}(n)\}_{n=0}^{\kappa}} \quad \frac{1}{2}\left\|\mathbf{F}(0)-\boldsymbol{\mu}_{\mathbf{F}_{\mathrm{t}}(0)}\right\|^2_{\mathbf{P}_{\mathrm{t}}^{-1}(0)}+\frac{1}{2}\sum_{n=0}^{\kappa-1}\|\mathbf{N}(n)\|^2_{\mathbf{L}_{\mathrm{t}}^{-1}(n)}$$

$$+\frac{1}{2}\sum_{n=0}^{\kappa}\|\boldsymbol{\mathcal{E}}[n,\boldsymbol{\Theta}(n)]\|^2_{\mathbf{S}_{\mathrm{t}}^{-1}(n)} \tag{5.45a}$$

$$\text{subject to} \quad \mathbf{F}(n+1)=\alpha_{\mathrm{t}}\mathbf{F}(n)+\mathbf{N}(n). \tag{5.45b}$$

We will now show that the ICs of the frequency-domain problem in (5.45) correspond to those derived for the frame-based time-domain problem in (5.44). To this end, we introduce some suitable variable transformations in the optimization problem; namely, the different frequency-domain variables are related to their time-domain counterparts using different constraint matrices

defined in tables 5.1 and 5.2 [146], as follows:

$$\mathbf{f}(n) = \mathbf{F}_R^{-1} \, (\mathbf{G}_{M\times R}^{10})^{\mathrm{H}} \, \mathbf{F}(n) \tag{5.46}$$

$$\mathbf{F}(n) = \mathbf{G}_{M\times R}^{10} \, \mathbf{F}_R \, \mathbf{f}(n) \tag{5.47}$$

$$\mathbf{j}(n) = \mathbf{F}_{n_J}^{-1} \, (\mathbf{G}_{M\times n_J}^{10})^{\mathrm{H}} \mathbf{J}(n) \tag{5.48}$$

$$\mathbf{J}(n) = \mathbf{G}_{M\times n_J}^{10} \, \mathbf{F}_{n_J} \, \mathbf{j}(n) \tag{5.49}$$

$$\mathbf{\Theta}(n) = [\mathbf{F}^{\mathrm{T}}(n) \quad \mathbf{J}^{\mathrm{T}}(n)]^{\mathrm{T}} \tag{5.50}$$

$$\mathbf{n}(n) = \mathbf{F}_R^{-1} \, (\mathbf{G}_{M\times R}^{10})^{\mathrm{H}} \, \mathbf{N}(n), \tag{5.51}$$

$$\mathbf{N}(n) = \mathbf{G}_{M\times R}^{10} \, \mathbf{F}_R \, \mathbf{n}(n). \tag{5.52}$$

Here we have assumed that $\mathbf{f}(n)$, $\mathbf{j}(n)$ and $\mathbf{n}(n)$ have lengths R, n_J and R, respectively, while all the frequency-domain variables have length M.

In addition to the transformations (5.47), (5.49) and (5.52), the four steps necessary to rewrite (5.45) as (5.44) are the following:

- The frequency-domain and time-domain feedback path model and process noise are related via (5.47) and (5.52); substituting (5.47) and (5.52) in (5.45b) and premultiplying with $\mathbf{F}_R^{-1} \, (\mathbf{G}_{M\times R}^{10})^{\mathrm{H}}$ leads to (5.44b).
- The first term of (5.45a) can be rewritten in terms of $\mathbf{f}(0)$ and $\mathbf{\Pi}_{\mathrm{t}}^{-1}(0)$ using (5.47) and the following relation:

$$(\mathbf{G}_{M\times R}^{10})^{\mathrm{H}} \, \mathbf{P}_{\mathrm{t}}^{-1}(0) = \mathbf{F}_R \, \mathbf{\Pi}_{\mathrm{t}}^{-1}(0) \mathbf{F}_R^{-1} \, (\mathbf{G}_{M\times R}^{10})^{\mathrm{H}}, \tag{5.53}$$

  resulting in the first term of (5.44a).
- The second term of (5.45a) can be rewritten in terms of $\mathbf{n}(n)$ and $\mathbf{\Lambda}_{\mathrm{t}}^{-1}(n)$ using (5.52) and the following relation:

$$(\mathbf{G}_{M\times R}^{10})^{\mathrm{H}} \, \mathbf{L}_{\mathrm{t}}^{-1}(n) = \mathbf{F}_R \, \mathbf{\Lambda}_{\mathrm{t}}^{-1}(n) \mathbf{F}_R^{-1} \, (\mathbf{G}_{M\times R}^{10})^{\mathrm{H}}, \tag{5.54}$$

  resulting in the second term of (5.44a).
- The third term of (5.45a) can be rewritten in terms of $\boldsymbol{\varepsilon}[n, \boldsymbol{\theta}(n)]$ and $\mathbf{\Sigma}_{\mathrm{t}}^{-1}(n)$, using the following relations:

$$\boldsymbol{\mathcal{E}}[n, \mathbf{\Theta}(n)] = (\mathbf{G}_{R\times M}^{01})^{\mathrm{H}} \mathbf{F}_R \, \boldsymbol{\varepsilon}[n, \boldsymbol{\theta}(n)] \tag{5.55}$$

$$\mathbf{G}_{R\times M}^{01} \, \mathbf{S}_{\mathrm{t}}^{-1}(n) = \mathbf{F}_R \, \mathbf{\Sigma}_{\mathrm{t}}^{-1}(n) \mathbf{F}_R^{-1} \, \mathbf{G}_{R\times M}^{10}, \tag{5.56}$$

  resulting in the third term of (5.44a).

Overall, the transformations (5.47), (5.49) and (5.52) to (5.56) can be used to link the frequency-domain problem and solutions in (5.45) to the frame-based time-domain problem and solutions in (5.44). Thus the ICs C1 to C3 and the considerations drawn for the frame-based time-domain problem (5.44) hold for the frequency-domain problem (5.45), too.

5.5 PEM-Based Partitioned-Block Frequency Domain Kalman Filter

Even though the use of FDAF algorithms has been shown to be beneficial compared to time-domain algorithms for several aspects, an important issue that might limit the applicability of FDAF algorithms is the use of excessive filter lengths. High-order filters, motivated by long echo or feedback paths and/or high sampling frequencies, can lead to algorithmic noise [160]. Additionally, high-order filters increase the algorithmic delay, potentially making real-time solutions unfeasible. Even though the feedback path of a HA is usually relatively short, the very low delay requirements generally make FDAF algorithms unsuitable for HA applications.

A way to overcome this problem involves the use of a PB structure, i. e. the so-called partitioned-block frequency-domain adaptive filter (PBFDAF) [115, 155]. The PBFDAF requires the division of the feedback path model into P partitions of length $L \leq R = n_F$, thus allowing the lowering of the algorithmic delay from $2n_F - 1$ to $2L - 1$ [84]. The PBFDAF has been successfully applied in both AEC [156, 114] and AFC [84, 165, 166]. Specifically, for AFC, the PBFDAF has been combined with a PEM-based prewhitening filter, giving rise to the PEM-PBFDAF [84, 144]. More recently, a state-space version of the PBFDAF algorithm for AEC has been proposed [160, 158], which will be referred to as partitioned-block frequency-domain Kalman filter (PBFDKF) in the following.

In this section, we propose a modified version of the PBFDKF including the same decorrelation stage employed for the FDKF presented in section 5.3, by means of a PEM-based prewhitening, i. e. the PEM-PBFDKF.

We first define the partitioned version of the time-domain M-samples loudspeaker signal and the L-samples true feedback path, at frame κ for block $p = 0, \ldots, P-1$, as follows

$$\mathbf{u}_p(\kappa) = [u(\kappa R - pL - M + 1) \quad \ldots \quad u(\kappa R - pL)]^{\mathrm{T}} \tag{5.57}$$

$$\mathbf{f}_{\mathrm{t},p}(\kappa) = [f_{\mathrm{t}}(pL, \kappa) \quad \ldots \quad f_{\mathrm{t}}(pL + L - 1, \kappa)]^{\mathrm{T}}\,. \tag{5.58}$$

The partitioned signal vectors can be defined in a similar way to the non-partitioned ones, adding the specific block index; in the following definitions, we are always considering the time frame κ for block p. The time- and frequency-domain version of the prefiltered loudspeaker signal can be defined as follows:

$$\mathbf{u}_{\mathrm{J_t},p}(\kappa) = [u_{\mathrm{J_t}}(\kappa R - pL - M + 1) \quad \dots \quad u_{\mathrm{J_t}}(\kappa R - pL)]^{\mathrm{T}} \tag{5.59}$$

$$\mathbf{U}_{\mathrm{J_t},p}(\kappa) = \mathrm{diag}\left\{\mathbf{F}_M \, \mathbf{u}_{\mathrm{J_t},p}(\kappa)\right\}, \tag{5.60}$$

with the constraint $M \geq R + L - 1$, to ensure proper operations [84]. The PBFD representation of the true feedback path can be defined, similarly to (5.18a), as follows:

$$\mathbf{F}_{\mathrm{t},p}(\kappa) = \mathbf{G}^{10}_{M\times L} \, \mathbf{F}_L \, \mathbf{f}_{\mathrm{t},p}(\kappa). \tag{5.61}$$

Finally, the time- and frequency-domain versions of the prefiltered microphone signal and source signal frame can be defined as follows, with $V = M - L$:

$$\mathbf{y}_{\mathrm{J_t}}(\kappa) = [y_{\mathrm{J_t}}(\kappa R - V + 1) \quad \dots \quad y_{\mathrm{J_t}}(\kappa R)]^{\mathrm{T}} \tag{5.62}$$

$$\mathbf{Y}_{\mathrm{J_t}}(\kappa) = (\mathbf{G}^{01}_{V\times M})^{\mathrm{H}} \mathbf{F}_V \, \mathbf{y}_{\mathrm{J_t}}(\kappa) \tag{5.63}$$

$$\mathbf{e}(\kappa) = [e(\kappa R - V + 1) \quad \dots \quad e(\kappa R)]^{\mathrm{T}} \tag{5.64}$$

$$\mathbf{E}(\kappa) = (\mathbf{G}^{01}_{V\times M})^{\mathrm{H}} \mathbf{F}_V \, \mathbf{e}(\kappa). \tag{5.65}$$

As in the non-partitioned case, one of the $M - L + 1 = V + 1$ samples from the fast frequency-domain linear convolution is dropped to simplify the notation, resulting in a V-samples length for both $\mathbf{y}_{\mathrm{J_t}}(\kappa)$ and $\mathbf{e}(\kappa)$ [158]. A common choice [146, 155, 158] for the parameters is $L = V = M/2$, recalling that now $L = n_F/P$. The frame shift R and the signal block length V can be related via $R = V/\gamma$, where γ is the overlapping factor, usually chosen to be an integer [155].

The resulting PB state-space model is the following:

$$\mathbf{F}_{\mathrm{t},p}(\kappa + 1) = \alpha_{\mathrm{t}} \mathbf{F}_{\mathrm{t},p}(\kappa) + \mathbf{N}_{\mathrm{t},p}(\kappa), \qquad p = 0, \dots, P - 1 \tag{5.66a}$$

$$\mathbf{Y}_{\mathrm{J_t}}(\kappa) = \sum_{p=0}^{P-1} \mathbf{C}_{\mathrm{J_t},p}(\kappa) \mathbf{F}_{\mathrm{t},p}(\kappa) + \mathbf{E}(\kappa) \tag{5.66b}$$

where $\mathbf{C}_{\mathrm{J_t},p}(\kappa) = \mathbf{G}^{01}_{M\times M} \, \mathbf{U}_{\mathrm{J_t},p}(\kappa)$ is used to linearize the pth circular convolution between the partitions $\mathbf{u}_{\mathrm{J_t},p}(\kappa)$ and $\mathbf{f}_{\mathrm{t},p}(\kappa)$, similarly to the non-partitioned case. The partitioning of the state equation (5.66a) requires the

definition of $\mathbf{N}_{\mathrm{t},p}(\kappa)$, i. e. the process noise for the pth partition; the transition factor α_t, however, is still partition invariant [160, 158].

As for the non-partitioned case, we can apply Kalman filter to the model defined in (5.66) to obtain the linear MMSE estimate of the state $\mathbf{F}_{\mathrm{t},p}(\kappa)$; a set of equations very similar to (5.24) can be written as follows, for each partition $p = 0, \ldots, P-1$:

$$\mathbf{K}_p(\kappa) = \mathbf{P}_p(\kappa)\mathbf{C}^\mathrm{H}_{\mathrm{J_t},p}(\kappa)\left[\mathbf{C}_{\mathrm{J_t},p}(\kappa)\mathbf{P}_p(\kappa)\mathbf{C}^\mathrm{H}_{\mathrm{J_t},p}(\kappa) + \mathbf{\Psi}_{\mathrm{EE}}(\kappa)\right]^{-1} \tag{5.67a}$$

$$\hat{\mathbf{F}}^+_p(\kappa) = \hat{\mathbf{F}}_p(\kappa) + \mathbf{K}_p(\kappa)\left[\mathbf{Y}_{\mathrm{J_t}}(\kappa) - \sum_{p=0}^{P-1}\mathbf{C}_{\mathrm{J_t},p}(\kappa)\hat{\mathbf{F}}_p(\kappa)\right] \tag{5.67b}$$

$$\mathbf{P}^+_p(\kappa) = \left[\mathbf{I}_M - \mathbf{K}_p(\kappa)\mathbf{C}_{\hat{\mathrm{J}},p}(\kappa)\right]\mathbf{P}_p(\kappa) \tag{5.67c}$$

$$\hat{\mathbf{F}}_p(\kappa+1) = \alpha_\mathrm{t}\hat{\mathbf{F}}^+_p(\kappa) \tag{5.67d}$$

$$\mathbf{P}_p(\kappa+1) = \alpha_\mathrm{t}^2\mathbf{P}^+_p(\kappa) + \mathbf{\Psi}_{\mathrm{N_tN_t},p}(\kappa). \tag{5.67e}$$

The simplified form of (5.67) relies on similar approximations as in the non-partitioned case, where R is replaced by V, i. e. $\mathbf{G}^{01}_{M\times M} \approx (V/M)\mathbf{I}_M$ and $\mathbf{G}^{01}_{M\times M}\,\mathbf{\Delta}(\mathbf{G}^{01}_{M\times M})^\mathrm{H} \approx (V/M)\mathbf{\Delta}$, if $\mathbf{\Delta}$ is a diagonal matrix [160, 158]. This allows to adapt the approximations listed in section 5.3 accordingly, yielding the diagonalized version of the algorithm, i. e. the PEM-PBFDKF:

$$\mathbf{K}_p(\kappa) = \mathbf{P}_p(\kappa)\mathbf{U}^\mathrm{H}_{\hat{\mathrm{J}},p}(\kappa)\left[\mathbf{U}_{\hat{\mathrm{J}},p}(\kappa)\mathbf{P}_p(\kappa)\mathbf{U}^\mathrm{H}_{\hat{\mathrm{J}},p}(\kappa) + M\ \mathrm{diag}\{\mathbf{\Phi}_{\hat{\mathrm{E}}\hat{\mathrm{E}}}(\kappa)\}\right]^{-1} \tag{5.68a}$$

$$\hat{\mathbf{F}}^+_p(\kappa) = \hat{\mathbf{F}}_p(\kappa) + \mathbf{G}^{10}_{M\times M}\,\mathbf{K}_p(\kappa)\boldsymbol{\mathcal{E}}[\kappa, \hat{\mathbf{\Theta}}(\kappa)] \tag{5.68b}$$

$$\mathbf{P}^+_p(\kappa) = \left[\mathbf{I}_M - \frac{V}{M}\mathbf{K}_p(\kappa)\mathbf{U}_{\hat{\mathrm{J}},p}(\kappa)\right]\mathbf{P}_p(\kappa) \tag{5.68c}$$

$$\hat{\mathbf{F}}_p(\kappa+1) = \alpha\hat{\mathbf{F}}^+_p(\kappa) \tag{5.68d}$$

$$\mathbf{P}_p(\kappa+1) = \alpha^2\mathbf{P}^+_p(\kappa) + M\ \mathrm{diag}\{\mathbf{\Phi}_{\hat{\mathrm{N}}\hat{\mathrm{N}},p}(\kappa)\}. \tag{5.68e}$$

A summary of the PEM-PBFDKF is given in algorithm 3, where the explicit calculations of $\mathbf{\Phi}_{\hat{\mathrm{E}}\hat{\mathrm{E}}}(\kappa)$ and $\mathbf{\Phi}_{\hat{\mathrm{N}}\hat{\mathrm{N}},\mathrm{p}}(\kappa)$ follow the procedures described in [157, 160, 158]; in particular, $\mathbf{\Phi}_{\hat{\mathrm{N}}\hat{\mathrm{N}},\mathrm{p}}(\kappa)$ is estimated using a first-order recursive filter,

Algorithm 3 PEM-based partitioned-block frequency-domain Kalman filter (PEM-PBFDKF). The PEM-FDKF can be obtained by setting $P = 1$.

$\mathbf{P}_p(0) = \sigma\mathbf{I}_M$, $\hat{\mathbf{F}}_p(\kappa) = \mathbf{\Phi}_{\hat{\mathrm{N}}\hat{\mathrm{N}},p}(0) = \tilde{\mathbf{\Phi}}_{\hat{\mathrm{F}}\hat{\mathrm{F}},p}(0) = \mathbf{0}_{M\times 1}$, $\hat{\mathbf{j}}(0) = \begin{bmatrix}1 & \mathbf{0}_{n_J-1\times 1}\end{bmatrix}^{\mathrm{T}}$
for $\kappa = 0, 1, 2, \ldots$ **do**
  for $p = 0, \ldots, P-1$ **do**
    $\mathbf{U}_p(\kappa) = \mathrm{diag}\left\{\mathbf{F}_M\, \mathbf{u}_p(\kappa)\right\}$,
  end for
  $\hat{\mathbf{y}}(\kappa) = \mathbf{W}^{01}_{V\times M}\, \mathbf{F}_M^{-1}\, \sum_{p=0}^{P-1} \mathbf{U}_p(\kappa)\hat{\mathbf{F}}_p(\kappa)$
  $\mathbf{d}(\kappa) = \mathbf{y}(\kappa) - \hat{\mathbf{y}}(\kappa)$
  $u_{\hat{\mathrm{j}}}(\kappa R - pL - i) = \hat{J}(q,\kappa)u(\kappa R - pL - i)$
    $i = 0, \ldots, M-1,\ \ p = 0, \ldots, P-1$
  $y_{\hat{\mathrm{j}}}(\kappa R - i) = \hat{J}(q,\kappa)y(\kappa R - i), \quad i = 0, \ldots, V-1$
  for $p = 0, \ldots, P-1$ **do**
    $\mathbf{U}_{\hat{\mathrm{j}},p}(\kappa) = \mathrm{diag}\left\{\mathbf{F}_M\, \mathbf{u}_{\hat{\mathrm{j}},p}(\kappa)\right\}$,
  end for
  $\bar{\mathbf{y}}_{\hat{\mathrm{j}}}(\kappa) = \mathbf{W}^{01}_{V\times M}\, \mathbf{F}_M^{-1}\, \sum_{p=0}^{P-1} \mathbf{U}_{\hat{\mathrm{j}},p}(\kappa)\hat{\mathbf{F}}_p(\kappa)$
  $\boldsymbol{\varepsilon}[\kappa, \hat{\boldsymbol{\theta}}(\kappa)] = \mathbf{y}_{\hat{\mathrm{j}}}(\kappa) - \bar{\mathbf{y}}_{\hat{\mathrm{j}}}(\kappa)$
  $\boldsymbol{\mathcal{E}}[\kappa, \hat{\mathbf{\Theta}}(\kappa)] = \mathbf{F}_M\, (\mathbf{W}^{01}_{V\times M})^{\mathrm{H}}\boldsymbol{\varepsilon}[\kappa, \hat{\boldsymbol{\theta}}(\kappa)]$
  $\mathbf{\Phi}_{\hat{\mathcal{E}}\hat{\mathcal{E}}}(\kappa) = \mathrm{diag}\{\boldsymbol{\mathcal{E}}[\kappa, \hat{\mathbf{\Theta}}(\kappa)]\boldsymbol{\mathcal{E}}^{\mathrm{T}}[\kappa, \hat{\mathbf{\Theta}}(\kappa)]\}/V$
  $\mathbf{\Phi}_{\hat{\mathrm{E}}\hat{\mathrm{E}}}(\kappa) = \mathbf{\Phi}_{\hat{\mathcal{E}}\hat{\mathcal{E}}}(\kappa) - \frac{\mathrm{diag}\{\sum_{p=0}^{P-1} \mathbf{U}_{\hat{\mathrm{j}},p}(\kappa)\mathbf{P}_p(\kappa)\mathbf{U}^{\mathrm{H}}_{\hat{\mathrm{j}},p}(\kappa)\}}{M}$
  Threshold $\mathbf{\Phi}_{\hat{\mathrm{E}}\hat{\mathrm{E}}}(\kappa)$ with $\hat{\sigma}^2_{\boldsymbol{\varepsilon}}(\kappa)\mathbf{1}_{M\times 1}$
  for $p = 0, \ldots, P-1$ **do**
    $\mathbf{K}_p(\kappa) = \mathbf{P}_p(\kappa)\mathbf{U}^{\mathrm{H}}_{\hat{\mathrm{j}},p}(\kappa)\Big[\mathbf{U}_{\hat{\mathrm{j}},p}(\kappa)\mathbf{P}_p(\kappa)\mathbf{U}^{\mathrm{H}}_{\hat{\mathrm{j}},p}(\kappa) + M\ \mathrm{diag}\{\mathbf{\Phi}_{\hat{\mathrm{E}}\hat{\mathrm{E}}}(\kappa)\}\Big]^{-1}$
    $\hat{\mathbf{F}}^{+}_p(\kappa) = \hat{\mathbf{F}}_p(\kappa) + \mathbf{G}^{10}_{M\times M}\, \mathbf{K}_p(\kappa)\boldsymbol{\mathcal{E}}[\kappa, \hat{\mathbf{\Theta}}(\kappa)]$
    $\hat{\mathbf{F}}_p(\kappa+1) = \alpha\hat{\mathbf{F}}^{+}_p(\kappa)$
    $\mathbf{P}^{+}_p(\kappa) = [\mathbf{I}_M\ - \frac{V}{M}\mathbf{K}_p(\kappa)\mathbf{U}_{\hat{\mathrm{j}},p}(\kappa)]\mathbf{P}_p(\kappa)$
    Threshold $\mathbf{P}^{+}_p(\kappa)$ with $\mathbf{0}_{M\times M}$
    $\mathbf{P}_p(\kappa+1) = \alpha^2\mathbf{P}^{+}_p(\kappa) + M\ \mathrm{diag}\{\mathbf{\Phi}_{\hat{\mathrm{N}}\hat{\mathrm{N}},p}(\kappa)\}$
    $\tilde{\mathbf{\Phi}}_{\hat{\mathrm{F}}\hat{\mathrm{F}},p}(\kappa+1) = \beta\tilde{\mathbf{\Phi}}_{\hat{\mathrm{F}}\hat{\mathrm{F}},p}(\kappa) + \frac{(1-\beta)\mathrm{diag}\{\hat{\mathbf{F}}_p(\kappa)\hat{\mathbf{F}}^{\mathrm{H}}_p(\kappa)\}}{M}$
    $\mathbf{\Phi}_{\hat{\mathrm{N}}\hat{\mathrm{N}},p}(\kappa+1) = (1-\alpha^2)\tilde{\mathbf{\Phi}}_{\hat{\mathrm{F}}\hat{\mathrm{F}},p}(\kappa+1)$
  end for
  $\hat{\mathbf{j}}(\kappa+1) = \texttt{Levinson} - \texttt{Durbin}\left([\mathbf{d}(\kappa); \mathbf{d}(\kappa-1)]\right)$
end for

with forgetting factor $\beta = 0.91$ [158]. Additionally, $\mathbf{1}_{M\times 1}$ is defined as an $M \times 1$ vector of ones.

Usually, in PBFDAF algorithms increasing the number of partitions reduces the convergence speed [167], since this increase results in smaller partitions, thus lowering the degree of decorrelation introduced when working in the frequency domain. However, this behavior is not always observed in the AFC case due to the closed-loop nature of the system. In a feedback scenario, if the system is effectively subject to or close to instability, the high power of the loudspeaker signal leads to a faster identification of the unknown feedback path and hence increases the convergence speed [82].

Finally, as pointed out by Buchner *et al.* [116, 168], PBFDAF algorithms relying on diagonal approximations require a stronger regularization than non-partitioned algorithms. For this purpose, in our implementation we introduce an additional thresholding operation in the calculation of $\mathbf{P}_p^+(\kappa)$, since the subtraction in (5.68c) may give rise to negative values in some frequency bins with low signal-to-noise ratio (SNR).

5.6 Computational Complexity

In this section, we provide a complexity analysis of the proposed PEM-PBFDKF algorithm, in comparison with the PEM-based partitioned-block frequency-domain adaptive filter (PEM-PBFDAF) [84], the FDAF [113], as well as the time-domain algorithms NLMS and PEM-AFROW [109], counting the number of per-output-sample real multiplications [113]. The following assumptions are made: a real multiplication and a real division have equal complexity; each length-M FFT/IFFT operation has a complexity of $D = M\log_2(M)$ multiplications [113]; the Levinsion-Durbin algorithm on a length-M signal vector has a complexity of $n_J^2 + (5+M)n_J + M$ multiplications. Table 5.3 lists the per-output-sample computational complexity. The normalization by R in the frequency-domain algorithms is only included to simplify the comparison; in reality, the system implementing the algorithms has a time equivalent to R samples to carry out a whole algorithm iteration.

The complexity of the different algorithms as a function of the prefilter order n_J is shown at the top of fig. 5.3. The results are obtained using $n_F = L = V = R = M/2 = 80/P$, $P = 1$ for the two non-partitioned algorithms (PEM-FDAF and PEM-FDKF), $P = \{2, 4\}$ for the two partitioned algorithms (PEM-PBFDAF and PEM-PBFDKF), and fixing the overlapping factor to $\gamma = 1$. A subscript is used to indicate the number of partitions and overlapping factor, respectively, when $P > 1$, e. g. PEM-PBFDAF$_{21}$ refers to the case with $P = 2$ and $\gamma = 1$.[1]

[1] For the sake of simplicity, the PB entries in table 5.3 include also the non-partitioned algorithms, i. e. PEM-FDAF and PEM-FDKF, using $P = 1$.

Table 5.3: Per-output-sample computational complexity and memory requirements of the compared algorithms. A numerical value is given in both cases for $n_F = L = V = R = M/2 = 80/P$, $n_J = 15$, and $P = \{1, 2, 4\}$.

Algorithm	Computational complexity	#	Memory requirements	#
NLMS	$3R+2$	322	$4R+7$	328
PEM-AFROW	$\frac{n_J^2+(M+3R+3)n_J+(1+R)M+2R(2R+3)}{R}$	568	$5R+n_H+15$	430
FDAF	$\frac{5D+13M}{R}$	99	$20M+R+4$	3284
PEM-PBFDAF	$\frac{(4P+3)D+3n_J^2+(2M+R+7)n_J+(2+17P)M+R}{R}$	226 | 307 | 457	$8M+3R+n_J+2(P-1)L+19PM+4$	4579 | 3899 | 3559
PEM-PBFDKF	$\frac{(4P+3)D+3n_J^2+(2M+R+7)n_J+(5+25P)M+R}{R}$	248 | 345 | 527	$10M+3R+n_J+2(P-1)L+31PM+5$	6820 | 5980 | 5560

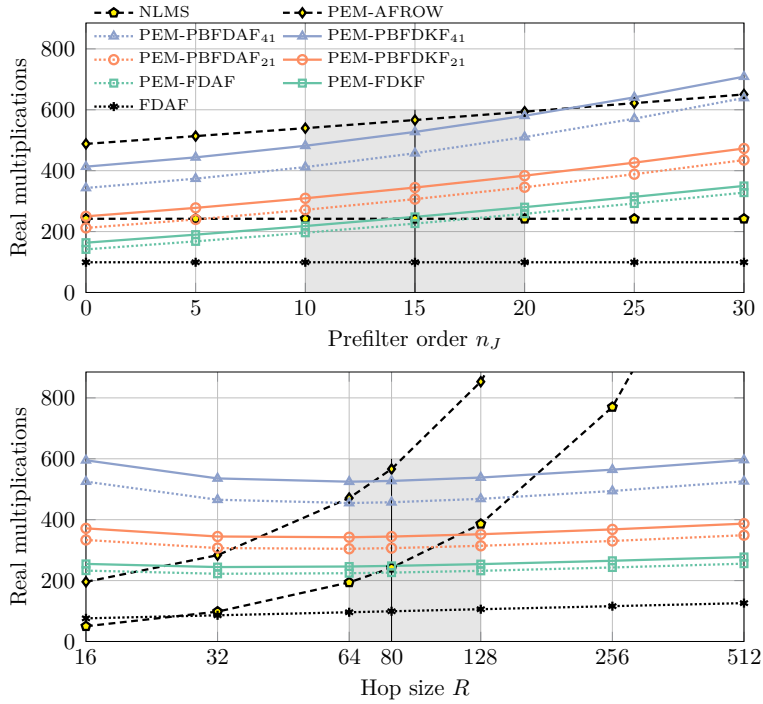


Figure 5.3: Per-output-sample computational complexity of the existing and proposed algorithms as a function of the prefilter order n_J and the hop size R (top and bottom, respectively).

For the PEM-AFROW, M and R are the window size and the hop size used to estimate the source signal model. The grayed part of the plots corresponds to values of n_J between 10 and 20, being common order values when using an AR source signal model for speech signals [84, 70, 144]. The value $n_J = 15$ is highlighted in the plot since it is used in the simulations presented in the following section. Using these values, the number of real multiplications for the different algorithms is also indicated in table 5.3, showing that the FDAF is the cheapest, while the PEM-AFROW is the most computationally expensive (even more expensive than the PB algorithms with four partitions). For the PEM-AFROW, a window length of $M = 160$ (10 ms at 16 kHz) might be too short as common values are in the order of 40-60 ms [109], but it was kept to simplify the comparison; by using longer windows, the complexity of the algorithm would be even higher.

The complexity of the different algorithms as a function of the hop size R is shown at the bottom of fig. 5.3. In this case, $n_J = 15$ and the other parameters

$n_F = L = V = R = M/2$ vary with R. The grayed part of the plots corresponds to values of R yielding tolerable latency values in a HA scenario [169]. As expected, one can see how the complexity of time-domain algorithms increases far quicker than the complexity of the frequency-domain algorithms, even within the grayed part.

We conclude the section presenting a worst-case estimate of the memory requirements for the considered algorithms (i. e. requiring the allocation of each variable, including temporary ones), which is also given in table 5.3 in terms of floating-point values to be allocated. Using $n_F = L = V = R = M/2 = 80/P$ and $n_J = 15$, the exact number of floating-point values to be allocated is also given. This shows that the requirements of frequency-domain algorithms are more than ten-fold those of time-domain algorithms and, as expected for the PB algorithms, the memory requirements decrease as P increases. Compared to the PEM-PBFDAF, the PEM-PBFDKF requires roughly 50 % more memory.

5.7 Simulations

In this section we provide simulation results in order to assess the performance of the proposed algorithms. The algorithms (proposed and baseline) are compared in terms of three measures, assessing the estimation error, the achievable amplification, and the sound quality.

The first measure is the misadjustment (Mis), defined as the dB level of the normalized distance between the true and estimated feedback path:

$$\mathrm{Mis}(\kappa) = 20 \log_{10} \frac{\|\mathbf{f}_\mathrm{r}(\kappa)\|}{\|\mathbf{f}_\mathrm{t}(\kappa)\|} \tag{5.69}$$

where $\mathbf{f}_\mathrm{r}(\kappa) = \mathbf{f}_\mathrm{t}(\kappa) - \hat{\mathbf{f}}(\kappa)$. Throughout the section, the vector containing the true feedback path coefficients will be referred to as acoustic impulse response (AIR).

The second measure is based on the so-called maximum stable gain (MSG), i. e. the maximum gain achievable at a given time without compromising the system stability; if the forward path $G(q, \kappa)$ is spectrally flat, the MSG is given as:

$$\mathrm{MSG}(\kappa) = -20 \log_{10} \left[\max_{l \in \mathcal{P}(\kappa)} |F_\mathrm{r}(\kappa, l)| \right] \tag{5.70}$$

where $\mathcal{P}(\kappa)$ is the set of frequencies satisfying the phase condition of the Nyquist stability criterion [46], and $F_\mathrm{r}(\kappa, l)$ is the lth element of $\mathbf{F}_\mathrm{r}(\kappa) = \mathbf{G}^{10}_{M \times R} \mathbf{F}_R \mathbf{f}_\mathrm{r}(\kappa)$.

The MSG is then normalized according to the maximum stable gain of the system when no feedback canceller is included, i. e. $K_{\mathrm{MSG}}(\kappa)$,

$$K_{\mathrm{MSG}}(\kappa) = -20\log_{10}\left[\max_{l\in\mathcal{P}(\kappa)}|F(\kappa,l)|\right] \tag{5.71}$$

resulting in the added stable gain (ASG):

$$\mathrm{ASG}(\kappa) = \mathrm{MSG}(\kappa) - K_{\mathrm{MSG}}(\kappa). \tag{5.72}$$

Finally, the sound quality is assessed by means of an objective measure called frequency-weighted log-spectral signal distortion (SD), a distance measure proven to correlate well with subjective evaluation of feedback cancellation algorithms [121]. The SD was calculated using the clean source signal $v(t)$ and the processed signal $d(t)$.

5.7.1 Stationary feedback path

For the first set of simulations, a stationary feedback path was employed. Such a scenario represents only a simplified situation, since a time variation in the feedback path can be caused by a multitude of different situations in real life, e. g. bringing a hand, or a phone close to the ear in a HA application. However, this simplified situation can be used to test the best-case scenario for the algorithms.

The following simulation results compare the performance of the PEM-FDKF and the PEM-PBFDKF for two different numbers of partitions, $P = \{2, 4\}$, and $\gamma = 1$ (PEM-PBFDKF$_{21}$ and PEM-PBFDKF$_{41}$, respectively), with three baseline algorithms. The chosen baseline algorithms are the PEM-FDAF [144] and the PEM-PBFDAF, based on the work by Spriet *et al.* [84], for two different numbers of partitions, $P = \{2, 4\}$, and $\gamma = 1$ (PEM-PBFDAF$_{21}$ and PEM-PBFDAF$_{41}$, respectively). The adaptive stepsize was calculated in a different way from what has been proposed in [84] and was defined as:

$$\mu_p(\kappa,l) = \mu_{\mathrm{fix}}\left[\delta + |U_{\hat{\mathrm{j}},p}(\kappa,l)|^2\right]^{-1}, \tag{5.73}$$

where μ_{fix} is a real, constant, and frequency-independent value, and δ is a regularization parameter. Some algorithm-specific parameters were kept fixed in the whole set of simulations as follows: the values of μ_{fix}, δ and $\mathbf{P}(0)$ are set as to guarantee a smooth convergence in the Mis curves in the first simulated case (cf. fig. 5.5), i. e. $\mu_{\mathrm{fix}} = 0.01$, $\delta = 5\times10^{-5}$ and $\mathbf{P}(0) = 1.6\times10^{-2}\,\mathbf{I}_M$, respectively; although, usually, when comparing adaptive filters performance one either tunes the parameters yielding similar lowest Mis or similar speed of

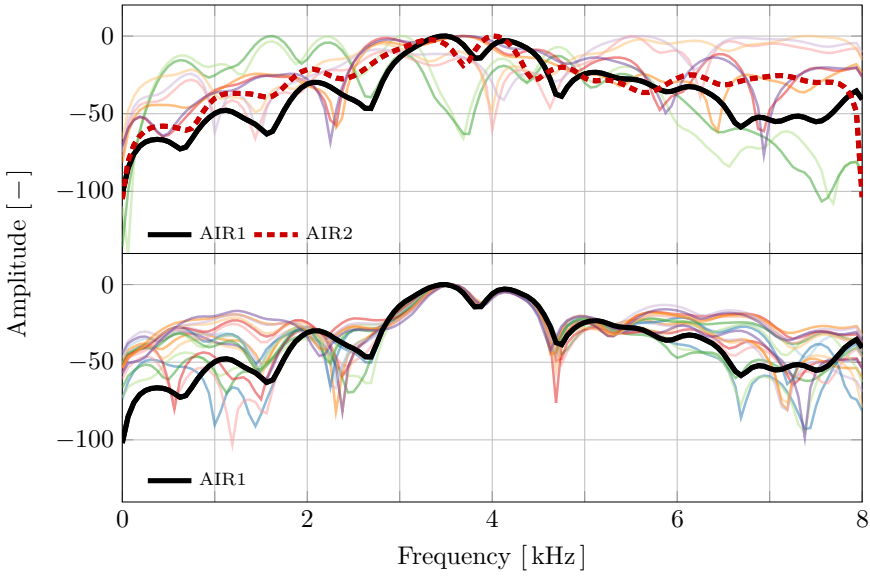


Figure 5.4: Top: Magnitude responses of the 10 different AIRs used in the simulations (cf. figs. 5.5, 5.6 and 5.8). Bottom: Magnitude responses of 20 AIRs obtained with (5.23a) every 1 s, starting from AIR1 (cf. fig. 5.7).

convergence, in this case, the proposed algorithms yield better performance in both ways.

Additionally, the transition factor is set to $\alpha = 0.99999$ [150], while the DFT length and the hop size are set to $M = 160/P$ and $L = R = V = 80/P$, respectively. The algorithms are tested in a simulated environment with the following characteristics: the transfer function $G(q, \kappa)$ is set to have constant gain $K_1 = K_{\text{MSG}} - 3\,\text{dB}$ and a delay of $d_G = 80$ samples; $F_{\text{t}}(q, \kappa) = F_{\text{t}}(q)$ is a measured feedback path from a commercial HA [84] and its normalized magnitude response is shown by the thick black line (AIR1) in fig. 5.4.

Figure 5.5 compares the algorithm performance, by means of Mis and ASG, when varying the input source signal $v(t)$: each one of 10 overlapping 20-s segments, taken from Track 05 of the Music for Archimedes database [170], of a recorded clean male speech signal resampled at $f_{\text{s}} = 16\,\text{kHz}$ corrupted by a 30 dB SNR additive white Gaussian noise are used as source signal $v(t)$. In fig. 5.5, and the following figures, the curve denotes the mean result of each algorithm while the shading corresponds to ± 1 standard deviations.

The results confirm the expected performance increase of the PEM-FDKF over

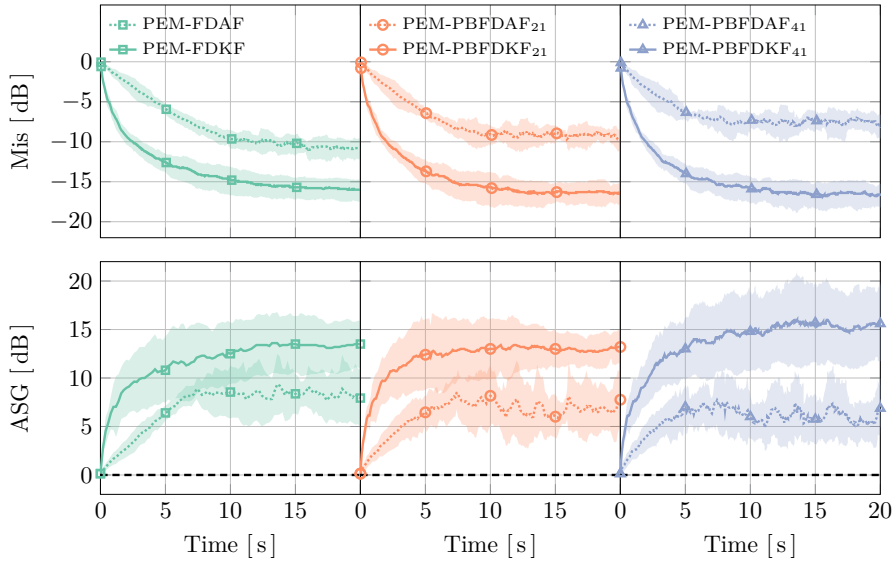


Figure 5.5: Mis and ASG performance over time of PEM-FDAF, PEM-FDKF, PEM-PBFDAF and PEM-PBFDKF using the constant forward path gain K_1, AIR1, and 10 different source signals.

PEM-FDAF [150]. Specifically, in fig. 5.5, similar convergence speeds and steady-state values are are observed in the Mis curves for the PEM-FDAF and PEM-PBFDAF, with a slight improvement in both the metrics as the number of partition increases. A similar situation characterizes the ASG values. For the PEM-FDKF and PEM-PBFDKF, similar Mis and ASG values are obtained as the number of partitions is increased; a higher ASG is obtained for the PEM-PBFDAF$_{41}$, even though a greater variance is observed. The performance obtained with the FDKF-based algorithms is better, both in terms of Mis and ASG, than the performance obtained with the FDAF-based algorithms, with improvements in the range from 5 to 10 dB.

Another parameter that might change the performance of the tested algorithms is the true feedback path included in the simulation. To simulate this variability we set up a similar simulation scenario as in the previously discussed case: 10 different feedback path AIRs (cf. top of fig. 5.4) were employed, while the source signal $v(t)$ was fixed to a single segment of the track used in the previous simulation; the other parameters were kept fixed. The results are shown in fig. 5.6 and provide a similar trend as in the previously discussed case regarding the performance improvements when using FDKF-based algorithms. The convergence curves are less smooth than in the previously discussed case and

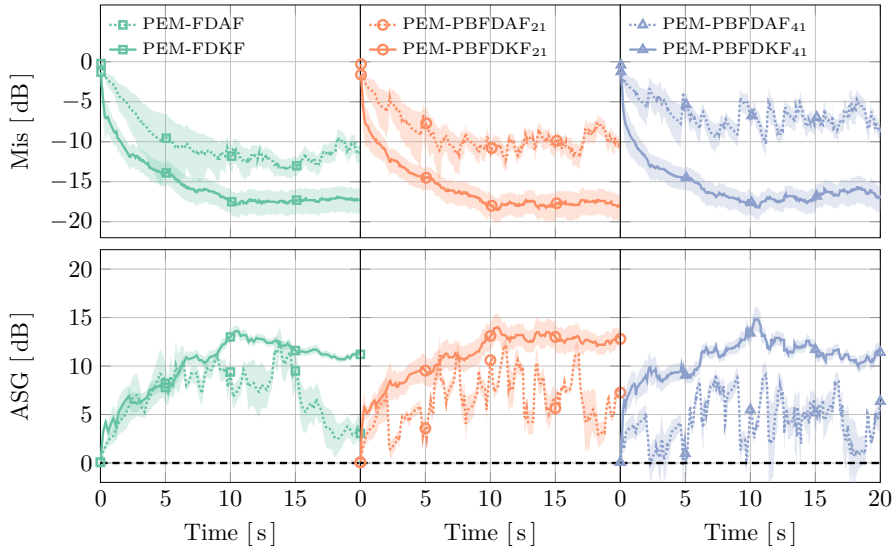


Figure 5.6: Mis and ASG performance over time of PEM-FDAF, PEM-FDKF, PEM-PBFDAF and PEM-PBFDKF using the constant forward path gain K_1 and 10 different AIRs.

the variance is lower in the single estimates, suggesting that both the smoothness and the variance of the curves is mainly depending on the source signal content, rather than on the AIR. Nevertheles, the FDKF-based algorithms yield much less erratic results than the FDAF-based algorithms; this is due to the need of a finer tuning of the FDAF-based algorithms parameters, e. g. μ_{fix} in (5.73), when dealing with different feedback paths, while the FDKF-based algorithms handle these differences more robustly.

The sound quality results, in terms of mean and maximum SD, for the first two simulation scenarios are shown in table 5.4 and labeled with the figure number showing the corresponding case. In both the scenarios the FDKF-based algorithms yield better performance and the increase of the number of partitions does not affect the results remarkably.

5.7.2 Non-stationary feedback path

For the second set of simulations, a non-stationary feedback path was employed. Both a smooth time evolution and an abrupt change were simulated. The first type of non-stationarity is given by the model described in (5.23a), where $\mathbf{F}(0)$

is AIR1, and the "true" simulated transition factor is $\alpha_t = 0.99999$. The process noise vector is generated as a complex normally-distributed random process with zero mean and variance set equal to a scaled version of the frequency-dependent variance of the 10 recorded AIRs shown at the top of fig. 5.4; this was done to avoid high variability at the beginning and at the end of the AIRs, in order to match more realistic AIR profiles. In this simulation, for the FDKF-based algorithms, both the transition factor and the frequency-dependent variance of the process noise were matching the "true" ones used in (5.23a). The magnitude responses of 20 AIRs obtained with (5.23a) every after 1 s of simulation are shown at the bottom of fig. 5.4. The results of the simulation are shown in fig. 5.7. The performance are worse than in the stationary case, but it should be noted that (5.23a) yields a strongly nonstationary behavior, i. e. a true feedback path changing every single frame. Despite the very fast dynamics of the "true" feedback path, the FDKF-based algorithms manage to provide a performance gap in the Mis between 2 to 5 dB. Similar conclusions hold for the ASG, with improvements of the FDKF-based algorithms over the FDAF-based algorithms being lower than in the stationary case.

The final simulation scenario considered in this section involves an abrupt feedback path change, from AIR1 to AIR2 (cf. fig. 5.4), at half of the simulation period. The results are shown in fig. 5.8. It can be noticed how the abrupt change causes a performance drop, which is naturally more evident for the FDKF-based algorithms since, after 10 s, these have already reached almost −15 dB Mis and 12 to 14 dB ASG, roughly 5 dB better than the performance measures of the FDAF-based algorithms at the same time. However, unlike the FDAF-based algorithms reconverging to similar levels, the FDKF-based algorithms behave differently after the abrupt change, showing an improved convergence speed as the number of partitions is increased. While the effect on the PEM-FDKF algorithm is expected, since the abrupt change causes a model mismatch in the state equation, the better performance of the PBFDKF algorithms might be due to the better tracking behavior due to the smaller block size of the PB implementations.

The sound quality results for the last two simulation scenarios, shown in table 5.4, confirm the findings from the first two. Also in this case, an increased number of partitions provides lower SD values than for the non-partitioned case, possibly due to the improved tracking behavior of the algorithm in case of a smaller block size.

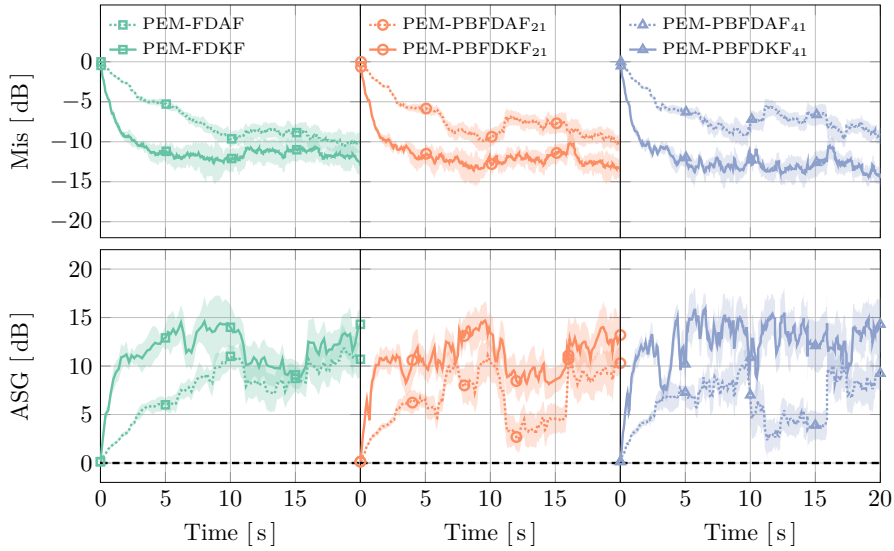


Figure 5.7: Mis and ASG performance over time of PEM-FDAF, PEM-FDKF, PEM-PBFDAF and PEM-PBFDKF using the constant forward path gain K_1, a smooth AIR transition (cf. bottom of fig. 5.4), simulated 10 times.

Table 5.4: Mean and maximum SD values for the different algorithms in the five different simulated scenarios.

	mean(SD) | max(SD)							
	fig. 5.5		fig. 5.6		fig. 5.7		fig. 5.8	
PEM-FDAF	0.93	2.77	0.85	2.79	1.05	2.97	1.26	3.67
PEM-PBFDAF$_{21}$	1.06	3.39	1.20	4.51	1.13	3.62	1.30	3.91
PEM-PBFDAF$_{41}$	1.26	3.81	1.67	5.57	1.27	3.92	1.44	4.40
PEM-FDKF	0.54	2.60	0.61	2.99	0.74	2.74	1.10	3.87
PEM-PBFDKF$_{21}$	0.48	2.47	0.54	2.80	0.66	2.26	0.88	3.69
PEM-PBFDKF$_{41}$	0.47	2.35	0.54	2.20	0.67	2.39	0.68	3.36

5.8 Conclusions

In this paper, we have presented a new AFC algorithm employing a PEM-based signal prewhitening in combination with a FDKF; additionally, we have derived the ICs to successfully apply such algorithm in a closed-loop system scenario.

We have also proposed an extension of the algorithm by means of a PB implementation, which makes the algorithm more appealing for use in systems

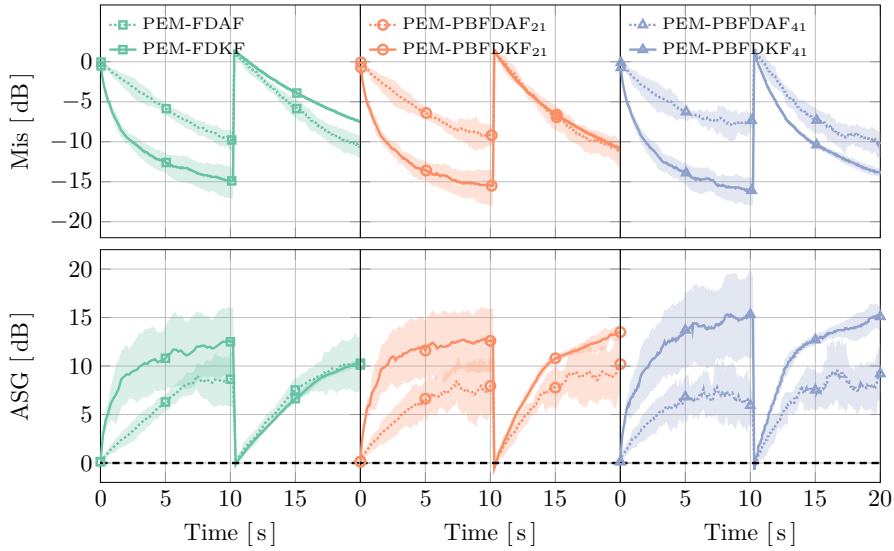


Figure 5.8: Mis and ASG performance over time of PEM-FDAF, PEM-FDKF, PEM-PBFDAF and PEM-PBFDKF using the constant forward path gain K_1, an abrupt AIR transition from AIR1 to AIR2, and 10 different source signals.

where large algorithmic delays are not tolerated, such as in HA applications.

A study of the computational complexity of the new algorithms has shown that, compared to existing frequency-domain algorithms, both the complexity increase and the memory requirements increase are limited.

Simulation results for different scenarios, including both stationary and non-stationary feedback paths, have shown the improved performance of the proposed algorithms in terms of different objective measures.

Acknowledgments

The authors would like to thank Prof. Piet Sommen and Dr. Fabian Kuech for their helpful discussions on PBFDAFs and Bang & Olufsen for providing the speech material.

Part III

Subjective and Objective Sound-Quality Evaluation

Chapter 6

Subjective and Objective Sound-Quality Evaluation of Adaptive Feedback Cancellation Algorithms

Subjective and Objective Sound-Quality Evaluation of Adaptive Feedback Cancellation Algorithms

Giuliano Bernardi, Toon van Waterschoot, Jan Wouters, and Marc Moonen

Submitted for publication to *IEEE/ACM Trans. Audio Speech Lang. Process.*, Nov. 2017.

The candidate's contributions as first author include: literature study, software implementation of the real-time algorithms, recording and post-processing of the test signals, co-design and supervision of the perceptual sound-quality evaluation, software implementation of the statistical analysis, software implementation and computer simulations for the evaluation of the objective scores, co-analysis of the subjective scores and of the subjective-to-objective correlations, co-formulation of the conclusions, text redaction and editing.

Abstract

Objective measures are widely used for the perceptual sound-quality evaluation of audio signal processing algorithms. Nevertheless, the use of subjective-evaluation measures remains relevant, in particular when application-specific objective evaluation measures are lacking.

In this paper, we present a perceptual sound-quality evaluation of different algorithms for adaptive feedback cancellation (AFC), with both speech and music signals. Three algorithms were compared: the block normalized least mean squares (BNLMS) algorithm, the prediction-error method (PEM)-based frequency-domain adaptive filter (PEM-FDAF), and the PEM-based frequency-domain Kalman filter (PEM-FDKF).

The perceptual evaluation results for the tested algorithms suggest that there is a strong statistical significance, and a corresponding large effect size, between the PEM-FDKF and the two remaining algorithms, when using speech signals. A smaller statistical significance, and a lower effect size, is reported when using music signals.

The perceptual evaluation results are then compared to the performance obtained with several objective measures. The correlation between subjective and objective scores shows that objective measures can be effectively used to predict the sound-quality degradation caused by acoustic feedback and AFC artifacts.

6.1 Introduction

Acoustic feedback is a common problem encountered in many acoustic systems, such as public address systems and hearing aids (HAs), when there exists an unwanted coupling between the loudspeaker and the microphone within the system. A system in a feedback configuration can give rise to an unstable behavior which, in acoustics terms, might translate into a spectrum of different phenomena affecting sound quality, such as signal distortion and howling [46]. This is why feedback cancellation is of considerable importance in such systems, especially in the case of HAs where small microphone-to-loudspeaker distances are involved.

In the past decades, many different algorithms have been developed to cope with acoustic feedback. A common class of acoustic feedback control algorithms includes the algorithms evaluated in this paper usually referred to as adaptive feedback cancellation (AFC) algorithms. The general idea in an AFC algorithm

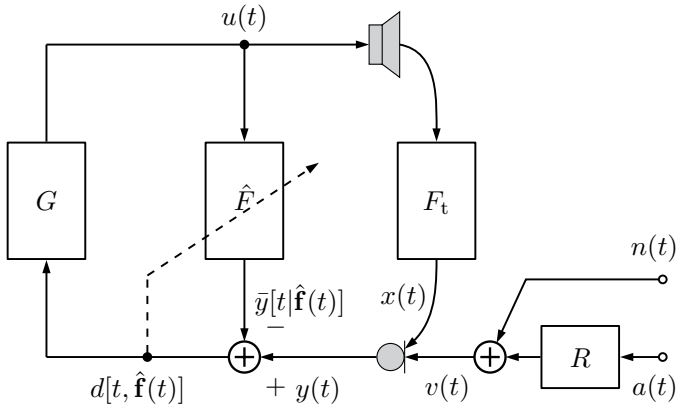


Figure 6.1: General AFC scenario.

is that of modelling the unknown true feedback path $F_{\mathrm{t}}(q,t)$, where the subscript $_{\mathrm{t}}$ stands for 'true', with an estimated adaptive filter $\hat{F}(q,t)$, where t and q^{-1} are the discrete time index and the delay operator, respectively, i. e. $q^{-k}u(t) = u(t-k)$. The loudspeaker signal $u(t)$ is then filtered by $\hat{F}(q,t)$ and the resulting output $\bar{y}[t|\hat{\mathbf{f}}(t)]$ is subtracted from the microphone signal $y(t)$, in order to remove the unwanted contribution of the feedback signal $x(t)$. The microphone signal $y(t)$ then approximates the source signal $v(t)$, modeled as the summation of the noise signal $n(t)$ and the anechoic signal $a(t)$ filtered with the room impulse response (RIR) $R(q,t)$, that is processed by the forward path $G(q,t)$ and provided to the patient through the loudspeaker, cf. fig. 6.1. The main problem faced by AFC algorithms is caused by the correlation between the source signal $v(t)$ and the loudspeaker signal $u(t)$, leading to a biased estimate of the feedback path [97]. In order to overcome this problem, many variations of the AFC algorithm involving different decorrelation approaches have been proposed, e. g. [102, 103, 104, 105, 106, 92].

A widely studied approach to provide a suitable decorrelation between the source signal $v(t)$ and the loudspeaker signal $u(t)$ includes the application of the prediction-error method (PEM) [107]. The resulting algorithm, the PEM-based adaptive feedback cancellation (PEM-AFC) [83], yields good convergence performance and limited perceptual distortions, when compared to algorithms employing other decorrelation approaches [46, 84]. Various modifications of the PEM-AFC have been developed by changing some of the underlying assumptions and/or implementation strategies; e. g., among others, the PEM-based adaptive filtering with row operations (PEM-AFROW) [93] makes the algorithm suitable for long feedback paths, the PEM-based frequency-domain adaptive filter (PEM-

FDAF) [84, 110] offers a frequency domain implementation of the feedback path adaptation, and the PEM-based frequency-domain Kalman filter (PEM-FDKF) [72] replaces the standard normalized least mean squares (NLMS) type adaptive filter with a Kalman filter.

Traditional performance measures for AFC algorithms are the misadjustment, i. e. the normalized error between the true and estimated feedback path, and the increase in the maximum stable gain (MSG), i. e. the additional gain maintaining the system stability upon the adaptive filter insertion [76]. These measures give information about the feedback system under investigation and are especially advantageous in simulated scenarios. A few recent studies, e. g. [118, 119, 120], have tackled the problem of objectively evaluating AFC algorithms in a more general way, that can be used to objectively evaluate the AFC algorithms of different commercial HAs. Not only did these studies include objective measures to evaluate instability (e. g. the MSG), but also objective measures evaluating oscillation and distortions.

This collection of different objective measures are helpful to evaluate several aspects related to the performance of a tested AFC algorithm, but are not readily interpretable in terms of perceptual sound quality [46]. Among different perceptually meaningful sound attributes, sound quality plays an important role in HA [171, 172]. Although sound quality can be heavily impacted by feedback and AFC artifacts, only a handful of sound-quality objective measures have been validated in a feedback scenario [118, 122]. Many other sound-quality objective measures were not specifically developed for feedback and AFC artifacts, and their typical application is that of assessing the sound-quality degradation caused by acoustic noise or by transmission channel effects (i. e. linear filtering) [173, 174]. The efficacy of these objective measures in a feedback/AFC scenario has not being investigated.

The goal of this paper is twofold. First, we present the results of a perceptual subjective sound-quality evaluation of three different AFC algorithms in order to validate the hypothesis of superior performance of the PEM-FDKF over the PEM-FDAF and the block normalized least mean squares (BNLMS) algorithm, as observed in simulations [72]. Second, the test signals used to subjectively evaluate sound quality, are tested using eight different existing objective measures, in order to evaluate the predictive capabilities of these objective measures.

The paper is organized as follows. In section 6.2, we describe different aspects of the current study, related to both the subjective and the objective evaluations. In section 6.3, we describe how the subjective evaluation has been carried out, focusing on the design details and choices of the experiment. In section 6.4, we describe the different objective measures used to perform the objective

evaluation. In section 6.5, we describe the performance criteria used to evaluate the predictive capabilities of the objective measures. In section 6.6, we describe the results of the subjective and the objective evaluations, as well as the correspondence between the two. Finally, the conclusions are drawn in section 6.7.

6.2 Materials and Methods

In this section, we give a detailed description of different aspects of the current study, i. e. the chosen AFC algorithms, the signal recordings and preprocessing, the employed speech and music material, and the post processing introduced to limit the annoyance of unstable loudspeaker signals.

6.2.1 Algorithms

Three AFC algorithms are considered in the current study: the BNLMS, the PEM-FDAF and the PEM-FDKF.

The BNLMS algorithm [112] is one of the simplest adaptive algorithms that can be used for AFC; other algorithms, such as affine projection (AP) and recursive least squares (RLS), exist but their increased complexity still represents a major hurdle in HA applications [46]. We use a frame version, instead of the standard NLMS algorithm, in order to simplify the comparison with the other two algorithms which, being frequency-domain implementations, also operate in a frame fashion. The BNLMS was implemented using the `LMS Filter` block in Simulink.

The PEM-FDAF [144] integrates the decorrelating properties of the PEM, a widely studied technique for AFC [83, 110, 86, 109, 93], into the frequency-domain adaptive filter (FDAF). An AFC algorithm based on the PEM assumes, in its simplest form, the source signal $v(t)$ to be the output of an autoregressive system, fed with a white noise input signal. In such a case, estimating the unknown source signal model and prefiltering the adaptive filter input and desired signal with the inverse signal model leads to an increased decorrelation between the source signal $v(t)$ and the loudspeaker signal $u(t)$, thus reducing the estimation bias.

Finally, the PEM-FDKF [72] integrates the PEM into the frequency-domain Kalman filter (FDKF) [151], and has been shown to provide superior performance over the PEM-FDAF [72]. The PEM-FDKF exploits the properties of the FDKF, leading to a faster convergence and to an optimal estimation of the

step-size parameter [151], simplifying the design choices one has to make in order to use the algorithm.

The three algorithms were implemented in Simulink Realtime and run on a Speedgoat Audio real-time target machine (`https://www.speedgoat.ch`). The sampling frequency was set to $f_\mathrm{s} = 16\,\mathrm{kHz}$ and the resulting sampling time $T_\mathrm{s} = 62.5\,\mu\mathrm{s}$ was well above the task execution time of the algorithms, being lower than $2\,\mu\mathrm{s}$.

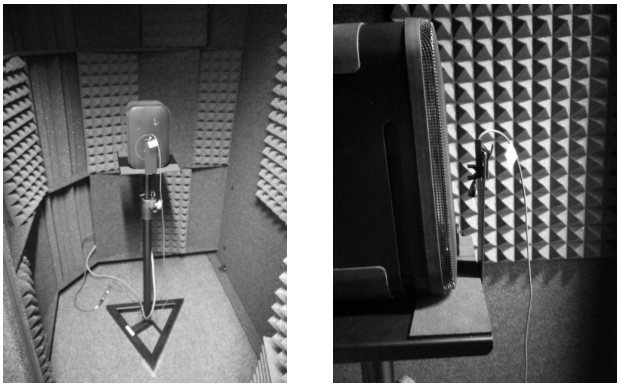
For all three algorithms, the frame size and the discrete Fourier transform (DFT) size were set to M, and the frame shift was set to R, where $M = 2R = 128$. The adaptive filter order was fixed to $n_{\hat{F}} = R - 1 = 63$, i. e. a 4 ms length.

Other algorithm-specific parameters were as follows: the stepsize of the BNLMS was either fixed to $\mu_\mathrm{anch} = 2 \times 10^{-2}$, when the algorithm was used to generate the anchor signal (cf. section 6.3.1), or to $\mu_\mathrm{BNLMS} = 2 \times 10^{-3}$. The adaptive stepsize of the PEM-FDAF was defined as in [72], with the fixed stepsize $\mu_\mathrm{FDAF} = 2 \times 10^{-2}$, and the regularization constant $\delta = 5 \times 10^{-5}$. The initial value of the error covariance matrix and the transition factor of the PEM-FDKF were set, respectively, to $\mathbf{P}(0) = 1 \times 10^{-4}\,\mathbf{I}_M$ and $\alpha = 0.999$. For more details about the PEM-FDAF and the PEM-FDKF we refer to [144] and [72]. These parameters were chosen such that each algorithm would remain stable for most part of all the source signals employed in the study. At the same time, the values were chosen not to cause the typical erratic behavior of a poorly-tuned adaptive filter.

Finally, to calculate the prefilter for the PEM-FDAF and the PEM-FDKF, the filter order was fixed to 15, cf. [72], while the estimation window length was set to 160 [84].

6.2.2 Signal recordings and preprocessing

The loudspeaker and the microphone were, respectively, a JBL CONTROL 1 PRO loudspeaker and an AKG CK 97-O microphone, placed in close proximity within a sound booth, see fig. 6.2. This was done to limit the length of the feedback path to similar values encountered in HA applications [46]. The forward path transfer function $G(q,t)$ was modeled using a gain and a delay stage. The gain was kept constant throughout each measurement, but a total of three different constant values were employed in different measurements. The three values were: $K_1 = K_\mathrm{MSG}$, $K_2 = K_\mathrm{MSG} + 10\,\mathrm{dB}$, and $K_3 = K_\mathrm{MSG} + 15\,\mathrm{dB}$, where K_MSG was the MSG of our setup when no AFC was included, cf. table 6.1. The delay of the forward path transfer function $G(q,t)$, $d_G = 80\,\mathrm{samples}$, was kept constant in all the measurements.



(a) From the front. (b) From the side.

Figure 6.2: Feedback scenario.

An automated script handled the recording of the different loudspeaker signals originating from the different recording combinations, i. e. varying speech/music signals, AFC algorithms, forward path gains and other algorithm parameters. The loudspeaker was driven by a signal coming from one of the Speedgoat output ports, passed through an RME Fireface UCX soundcard, a Samson SERVO 200 amplifier and, finally, reaching the loudspeaker using the feed-through of the sound booth. The loudspeaker signals were stored in the flash memory of the Speedgoat, and subsequently transferred to the connected PC running Simulink.

The microphone was connected, using the feed-through of the sound booth, to a EURORACK MX 3242X mixer, providing the necessary phantom power and, from there, to one of the Speedgoat input ports.

The source signal, $v(t)$ in fig. 6.1, was an anechoic speech or music signal, i. e. $a(t)$, filtered with a measured RIR and added to the microphone signal directly in the Simulink model. The signal-to-noise ratio (SNR) of the source signal $v(t)$ with respect to the background noise within the booth (stemming from the lab equipment outside the booth), i. e. $n(t)$, was fixed to 40 dB in all the recordings.

The frequency responses of the loudspeaker and microphone used in the setup, were objectively different than the frequency responses of the same components in a standard HA. In particular, the loudspeaker was much bigger and, as a consequence, had a much stronger low-frequency response. Additionally, the coupling of a HA receiver with the ear, provides a further filtering of

Table 6.1: Test signals IDs, descriptions and employed gains.

ID	Speech/music material and description	$G(q,t)$ [dB]
$\mathtt{S_{f1}}$	LIST\list01\{09, 08}	K_2
$\mathtt{S_{f2}}$	LIST\list01\{10, 05}	K_2
$\mathtt{S_{f3}}$	——	K_3
$\mathtt{S_{f4}}$	LIST list01\{03, 01, 04}	K_2
$\mathtt{S_{f5}}$	LIST\list04\{01, 04, 05}	K_2
$\mathtt{S_{f6}}$	LIST\list04\{09, 07}	K_2
$\mathtt{S_{m1}}$	LISTm\list01\{01, 02, 03}	K_2
$\mathtt{S_{m2}}$	LISTm\list03\{01, 02}	K_2
$\mathtt{S_{m3}}$	LISTm\list03\{03, 04}	K_2
$\mathtt{S_{m4}}$	LISTm\list03\{05, 06}	K_2
$\mathtt{S_{m5}}$	LISTm\list03\{07, 08}	K_2
$\mathtt{S_{m6}}$	——	K_3
$\mathtt{M_1}$	B. White - *You're the first, the last, my everything* - Rhythm and blues, disco	K_1
$\mathtt{M_2}$	B. Bragg - *A New England* - Folk rock	K_1
$\mathtt{M_3}$	——	K_2
$\mathtt{M_4}$	Los Mensajeros de las Libertad - *El Contrapunto* - Latin American folk	K_1
$\mathtt{M_5}$	Ska-P - *El Gato Lopez* - Ska, punk	K_1
$\mathtt{M_6}$	M. Ravel - *Bolero* - Classical	K_1
$\mathtt{M_7}$	W. A. Mozart - *Magic Flute Overture* - Classical	K_1
$\mathtt{M_8}$	J. S. Bach - *Partita for Violin No. 2* - Classical	K_1

the loudspeaker signal, that we did not account for in our scenario. Initial tests showed some problems related to this strong low-frequency behavior of the loudspeaker leading, e. g., to instabilities at low frequencies. The AFC algorithms tested in this study cannot cope with strong feedback at low frequencies, due to the limited number of adaptive filter coefficients [84]. Therefore, loudspeaker signals were filtered with an A-weighting filter [175], introducing a strong reduction of the frequencies below 500 Hz, in order to partially account for the discrepancies with a real HA scenario.

6.2.3 Speech and music material

Both speech and music signals were used in the study. The speech material was taken from the LIST [176] and LISTm [177] corpora, consisting of Dutch sentences uttered by a female and a male speaker, respectively. Each signal used in the study was obtained by merging either two or three utterances from

one of the lists of the corpora; this was done in order to have speech signals of length between 8 and 10 s, as suggested in the listening test standard [178], cf. section 6.3.1. Five signals were taken from each of the corpora and processed using gain K_2, i. e. $\mathtt{S}_{\mathtt{f1}}$ to $\mathtt{S}_{\mathtt{f5}}$ and $\mathtt{S}_{\mathtt{m1}}$ to $\mathtt{S}_{\mathtt{m5}}$; one of the five signals in both corpora (randomly chosen) was also processed using an additional forward path gain K_3, i. e. $\mathtt{S}_{\mathtt{f6}}$ and $\mathtt{S}_{\mathtt{m6}}$. In order to limit the number of stimuli and thus the duration of the listening test, only few signals were processed using gain K_3. Therefore, no statistical analysis can be done using the level as an independent variable. Nevertheless, the signals processed with gain K_3 were kept to offer a qualitative perspective on the results at higher gains. Further details of the utterances used to obtain the different speech signals employed in the sound-quality evaluation can be found in table 6.1.

The music material included a total of eight music signals, of different genres, drawn from different sources. Five music signals (including one repeated at two forward path gains), i. e. $\mathtt{M}_1$ to $\mathtt{M}_6$, were taken from the corpus associated with the Music Information Retrieval Evaluation eXchange (MIREX 2006) challenge [179]. Two signals, i. e. $\mathtt{M}_7$ to $\mathtt{M}_8$, were excerpts of classical music which have already been used to test AFC algorithms [46, 118]. All the signals were processed using a forward path gain K_1, and one of the signals was also processed using a forward path gain K_2. The choice of using a lower forward path gain (except for one recording) compared to the case of the speech material was motivated by literature results showing that AFC convergence is generally worse for music signals than for speech signals [46].

The implementations of the PEM-based algorithms used in this study rely on a low-order linear prediction model of the source signal and are, hence, not very good models for music signals [109]. Nevertheless, music signals were included given the increased evidence of hearing-impaired listeners enjoying music thanks to the advances in HA technology [180].

6.2.4 Signals post-processing

Due to the signal-dependent behavior of the AFC algorithms, guaranteeing stability of all the algorithms for all the different signals and the different parameters used in the listening test was found to be challenging. This was especially true for the anchor and the baseline algorithms (i. e. the BNLMS and the PEM-FDAF), which exhibited multiple instability periods during the recordings associated with different source signals. Some of the recordings were corrupted with artifacts that were too loud and, as a consequence, too disturbing for the sound-quality evaluation procedure.

Therefore, the test signals used in the listening test were obtained from the screened and post-processed loudspeaker signals, by means of targeted gain reductions in those time intervals corresponding to an unstable signal segment. Different gain reductions were applied, sometimes multiple times per each signal, using multiples of -5 dB. The gain reductions were applied in two speech signals processed by the BNLMS, in three speech signals processed by the PEM-FDAF, and in all the speech signals as well as in two music signals used as anchors.[1]

The post-processing was only applied to the anchor and to the baseline algorithms since processing the signal through these three algorithms led, to a different extent, to an unstable behavior in some cases; the PEM-FDKF did not exhibit any strong instability for the processed files and, thus, no post-processing was applied. Due to the way the reference signal was obtained, no feedback (and, naturally, no AFC) artifacts were present hence no post-processing was needed.

Although the inclusion of the post-processing stage does necessarily alter the outcome of the listening test, since it might have lead to higher scores for the baseline algorithms and the anchor than without any post-processing, we would like to point out that the main hypothesis we aim to test here is the superior performance of the PEM-FDKF. Therefore, we believe that our methodology is valid, even though it might lead to a more conservative assessment of the perceptual improvements of the PEM-FDKF over the baseline algorithms. Nevertheless, in section 6.6.1 we report the results of both the cases including and excluding the signals where post-processing was applied.

6.3 Subjective Evaluation

In this section, we give a detailed description of different aspects related to the subjective evaluation, i. e. the actual listening test and the chosen statistical analysis.

6.3.1 Listening test

The subjective sound-quality evaluation was performed using the Multi Stimulus Tests with Hidden Reference and Anchor (MUSHRA) [178]. The MUSHRA was preferred over other standards for sound-quality evaluation, e. g. [181], since

[1] A detailed summary of the gain reduction values applied to the different loudspeaker signals, including the time intervals when the gains were applied, is provided at the following link `https://lirias.kuleuven.be/bitstream/123456789/599139/5/fbMUSHRApp.csv`.

the predicted impairments were supposed to be, at least, intermediate. The implementation of the MUSHRA was adapted from `beaqleJS`, a Javascript and HTML5 based framework for subjective sound-quality evaluation [182]. The adaptation included the addition of the MUSHRA training phase and several small edits.

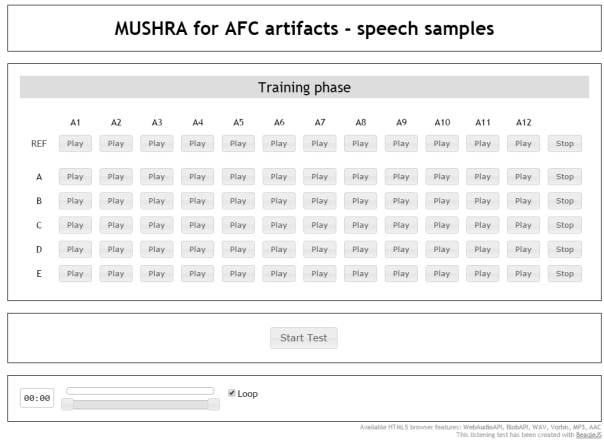
A total of 21 subjects were tested of which 12 were native Dutch speakers (the language of the speech corpora).[2] The subjects were 24 to 32 years old, and not all of them were experienced listeners, a recommended but not necessary condition for the test [178]. Prior to the actual listening test, the subjects were screened through an audiometric test run on a PC, using the APEX 3 program developed at KU Leuven ExpORL [183], connected to the RME Fireface UCX soundcard and a pair of standard audiometric headphones, the Sennheiser HDA200. The calibration of the setup was performed using the B&K 4153 artificial ear, including a B&K 4192 microphone. Upon testing, all the subjects were confirmed to be normal hearing.

The listening test was run on the same PC used for the audiological screening but, due to the small dimensions of the listening booth, an Apple iPad was used as an external user interface. The presentation level of the different test signals was 65 dB SPL, calibrated using the noise calibration signal of the LISTm corpus, and the signals were presented diotically via the same headphones used for the audiological screening. All the subjects acknowledged that the listening level was comfortable before starting the test.

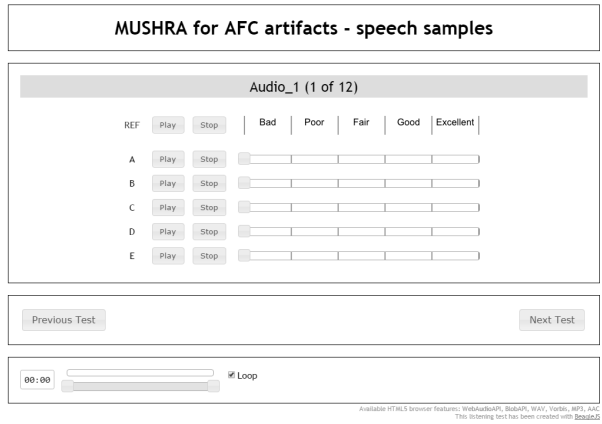
The listening test was divided in two main parts based on the kind of audio material employed in each part (speech and music, respectively). Each of these two parts was additionally divided into a training and a test phase. The training phase was conducted first, and provided the subjects with all the conditions that were subsequently presented during the test phase, see fig. 6.3a. This acclimation phase exposes the test subjects to all types of artifacts that are corrupting the test signals they were requested to assess in the actual test. The test phase consists of the presentation of the test signals, one at a time, processed through the different algorithms at test (cf. infra). The subjects must rate the sound quality associated with the different conditions compared to a reference (which is always accessible) on a scale from 0 to 100, with higher ratings indicating better sound quality, see fig. 6.3b.[3] The average duration of the listening test, including a brief description of the test (15 min) as well as training and test for both parts (62 min $\pm$ 11 min) was roughly 1 h 15 min.

[2] All participants gave written informed consent and ethical approval for the study was obtained from the Medical Ethics Committee UZ/KU Leuven.

[3] The numerical values were not displayed in the test user interface, but the subject were explicitly told that the grading bar ranged from 0 to 100.



(a) Screenshot of the training interface.



(b) Screenshot of the test interface.

Figure 6.3: Training and test interfaces of the listening test.

For each of the test signals, the subjects were presented with five conditions:

1. Three conditions corresponding to the loudspeaker signals obtained by running the three algorithms presented in section 6.2.1, i. e. the BNLMS, the PEM-FDAF, and the PEM-FDKF, possibly including a signal post-processing;

2. One hidden reference condition obtained by running the source signal through an "opened" feedback model, i. e. where the loudspeaker signal was logged right after the forward path stage but no signal was played

by the actual loudspeaker, in order to avoid any feedback into the microphone signal. In this way, the loudspeaker signal presented the same characteristics, in terms of forward path gain, SNR, microphone transfer function, and A-weighting as the other signals; however, no feedback (and, naturally, no AFC) artifacts were present.

3. One anchor specifically designed for this test, consisting of the loudspeaker signal obtained by running the BNLMS with non-optimal parameters (i. e. $\mu_{\text{anch}} = 2 \times 10^{-2}$), possibly including a signal post-processing. This value was chosen manually in order to generate significant feedback and AFC artifacts, e. g. whistling and entrainment [46], in all the loudspeaker signals. Therefore, the anchor should be easily recognized as a worst-case condition. No additional anchor obtained with the low-pass filtering described in [178] was used since the sound-quality degradation caused by a low-pass filter does not resemble the sound-quality degradation caused by feedback and AFC artifacts.

The subjects were not told which of the five conditions corresponded to which algorithms. They were, however, instructed that a hidden reference was included as one of the five conditions and that at least one of the conditions should be rated 100. When giving the test instructions, we stressed that the references would not be clean, natural-sounding test signals and, despite that, they should be associated with the best achievable sound quality in the considered experiment, i. e. equivalent to 100 in the rating scale.

6.3.2 Statistical analysis

The results, presented in section 6.6.1, from speech and music signals are analyzed separately. The chosen analysis technique is a mixed-design analysis of variance (ANOVA) [178], carried out with algorithm (excluding anchor and reference, i. e. the BNLMS, the PEM-FDAF, and the PEM-FDKF) and test signal as within-subject factors, while the nativeness of the test subject is included as a between-subject factor. The three variables (named 'effects' for consistency with the literature) will be indicated, respectively, as `AL`, `AU`, and `NA`. Besides the standard reported measures from an ANOVA analysis, i. e. degrees of freedom, F-values, and p-values p_A, also the effect size will be reported. The chosen effect size was the generalized eta-squared η_G^2, as it can be calculated along the ANOVA, and it was included given the increasing importance acknowledged to this metric for scientific reporting of statistical data [184, 185, 186]. The recommendations from [184] were used to interpret the effect of η_G^2: small if $\eta_G^2 \leq 0.02$, medium if $\eta_G^2 \approx 0.13$, and large if $\eta_G^2 \geq 0.26$.

In order to account for departures from sphericity, a Huynh-Feldt correction was applied to the p-values obtained from the ANOVA [178].

Along with the mixed-design ANOVA results, when the `AL` effect is found to be significant, also the results of a Tukey's HSD test on the considered effect is presented. This post hoc analysis can be used to find statistically significant differences, through the p-value p_T, between the results of the different algorithms and, hence, rank the algorithms in terms of overall sound quality. In this case the chosen effect size was the Cohen's d [184], using the pooled standard deviation between each pair as a normalization, whose values can be interpreted as follows: small if $0.2 < d < 0.3$, medium if $d \approx 0.5$, and large if $d > 0.8$ [184].

Whereas (mixed-design) ANOVAs are widely used to analyze MUSHRA data, especially within communities working in audio processing, based on their robustness to assumption violations [187, 188, 189], newer and more robust methods have been recently derived. Wilcox [190] collected several recent innovations in statistical methods to specifically deal with non normality and heteroscedasticity, as well as to achieve better control of type I errors. Most of these methods are now available in a number of R packages, such as `WRS`, `WRS2`, and `welchADF` [191, 192], among others. Therefore, an analysis was carried out with one of these packages, `welchADF` [193], implementing Johansen's general formulation of the Welch-James's test with approximate degrees of freedom with trimmed means, Winsorized variances and the use of bootsrap, as a way to verify the analysis carried out with the standard mixed-design ANOVA mentioned earlier. The results of the omnibus test are given in terms of the p-values p^{o}_{WJ}. The same statistics is applied as a post-hoc test to the `AL` effect, when the latter is found to be significant, and the results of the post-hoc test is returned in terms of Hochberg corrected p-values p^{p}_{WJ}.

6.4 Objective Evaluation

Different objective measures chosen from the literature were applied to the postprocessed files used for the listening test: the HA audio quality index (HAAQI) [194], the HA speech intelligibility index (HASPI) [195], the HA speech quality index (HASQI) [171, 122], the perceptual evaluation of audio quality (PEAQ) [196], the perceptual-model-based quality prediction method (PEMO-Q) [197], the perceptual evaluation of speech quality (PESQ) [174, 198], the frequency-weighted log-spectral signal distortion (SD) [118], and the short-time objective intelligibility (STOI) [199]. All these objective measures are intrusive, meaning that they rely on a reference signal to compute the degradation. Not

all of the objective measures are designed for sound-quality evaluation and, among those that were, most were not designed to evaluate the sound quality degradation caused feedback and AFC artifacts.

All the results obtained from the objective measures were normalized between 0 and 1, to simplify the comparison between different objective measures. Given that the SD does not provide a bounded value like the other objective measures, we mapped the maximum value found in the simulations to 1.

6.4.1 HAAQI, HASPI, and HASQI

HAAQI, HASPI and HASQI are three indexes developed by Kates and Arehart to predict audio quality, speech intelligibility and speech quality, respectively [194, 195, 171, 122]. These three indexes are all relying on an internal representation of the reference and processed signal with an auditory model, including a modelling stage to account for hearing impairment.

The auditory model was initially adopted by HASQI Version 1 [171] to predict speech quality variations due to noise, nonlinear distortions, and linear filtering in normal hearing (NH) and hearing impaired (HI) subjects. Subsequently, an extended version of the index, HASQI Version 2 [122], was introduced to cope with the degradations caused by frequency compression, noise suppression, speech vocoding, acoustic feedback and feedback cancellation, and speech combined with modulated noise. HASPI [195] adapts the auditory model used by HASQI in order to predict speech intelligibility. HAAQI [194] uses the HASQI auditory model with different parameters fitted to the data of an extensive music-quality rating experiment.

Although the three indexes can be used as predictors of quality and intelligibility for both NH and HI subjects, they are only used to predict NH scores, given the nature of our subjective data. In this paper, we used three MATLAB implementations of HAAQI, HASPI, and HASQI Version 2, respectively, kindly provided by Prof. Kates.

6.4.2 PEAQ

PEAQ is a perceptual measure standardized by the International Telecommunications Union (ITU) [196]. It relies on the combination of different auditory model outputs and has been developed to evaluate high-quality (i. e. low-distortion) digital codec systems. Two versions of PEAQ have been described in the literature: a basic and an advanced one. No implementation of PEAQ is freely available, but there exist at least three non-conformant open source

versions. In this paper, we used an implementation of the advanced PEAQ by Holters *et al.* [200].

6.4.3 PEMO-Q

PEMO-Q [197] returns the perceptual similarity measure (PSM) between a reference and a processed signal. It is based on a quantitative model of the "effective" auditory processing [201], that has been psychoacoustically validated in diverse tasks. For this reason, the PEMO-Q gives a higher prediction accuracy over PEAQ in several tasks [197]. In this paper, we used a MATLAB version of PEMO-Q from the PEASS toolkit [202].

6.4.4 PESQ

PESQ [174, 198, 203], as well as its wideband extensions [204], is part of an ITU Standard and is widely used to objectively assess the perceptual sound quality of speech. The PESQ score is obtained by processing both the reference and processed signal with an auditory model and combining different intermediate scores weighted from telephony data. In this paper, we used an implementation of the wideband PESQ built from the binaries provided by the ITU.

6.4.5 SD

SD is an easy-to-compute objective measure to assess speech quality [118, 46]; it relies on the short-term power spectral density (PSD) of the reference and the processed signals, $P_r(f,t)$ and $P_u(f,t)$ respectively, and on a psychoacoustically-motivated frequency weighting, i. e. w_{ERB} [118], based on auditory critical bands in the Nyquist interval:

$$SD(t) = \sqrt{\int_0^{f_s/2} w_{\mathrm{ERB}} \left(10 \log_{10} \frac{P_u(f,t)}{P_r(f,t)}\right)^2 \mathrm{d}f}. \tag{6.1}$$

The mean value of SD over the entire signal was used in the evaluation. SD has been used to evaluate sound quality in different papers analyzing AFC algorithms, e. g. [118, 119, 120, 46, 150], based on the good correlation results between SD and MSG found in [118]. However, its performance has not yet been validated with sound quality subjective data.

6.4.6 STOI

STOI is a correlation-based speech intelligibility measure that works on the temporal envelopes of short speech frames [199]. In recent years STOI has become extremely popular for the assessment of different kind of speech enhancement algorithms due to its simplicity and good performance. Therefore, it was included in our batch of objective measures, even though it is neither designed for sound quality evaluation nor for sound quality deterioration caused by feedback and AFC artifacts. In this paper, we used a MATLAB implementation of STOI from the webpage of the author (`http://www.ceestaal.nl/code/`).

6.5 Performance Criteria

Five commonly-used performance criteria, chosen based on the analysis in [205], were used to evaluate the predictive capabilities of the objective measures.

Three different correlation coefficients were computed. First, the Pearson correlation coefficient r was calculated to quantify linear relationships between objective scores O and subjective scores S. Second, the Spearman rank correlation coefficient r_{spear} was calculated to quantify how well O rank compared to S. Third, an additional Pearson correlation coefficient r_{sig} was calculated, assuming a possible nonlinear relationship between the scores O and S, in the following way:

1. A sigmoidal mapping was found between the pairs of scores O and S, i. e.

$$S = \frac{1}{1 + e^{-(\alpha_1 O + \alpha_2)}} \times 100\,\%, \tag{6.2}$$

   through nonlinear regression; this results in $\hat{\alpha}_1$ and $\hat{\alpha}_2$, estimates of α_1 and α_2, respectively;

2. The scores O were mapped to the sigmoidal-mapped objective scores O_{sig} using $\hat{\alpha}_1$ and $\hat{\alpha}_2$, ;

3. The Pearson correlation coefficient r_{sig} was calculated between O_{sig} and S.

Finally, the per-condition (i. e. per-algorithm) root-mean-square error (RMSE) and its extension, the so-called epsilon-insensitive root-mean-square error (ε-RMSE), were used. The ε-RMSE extends the RMSE in the sense that it assesses

the prediction error between O_{sig} and S by taking into account the variability of S, as recommended in the ITU-T Rec. P.863 [206]. The decision of including also the RMSE, along with the ε-RMSE, was made since, in some cases, a value of ε-RMSE $= 0$ was obtained. This was due to the high variability of the subjective data and the specific formulation of the ε-RMSE.

After computing the five performance criteria for each of the objective measures, a test to assess the statistical indifference among pairs of results was performed. The chosen statistical test for the correlation-based criteria was the Fisher transformation z-test [205]; for the RMSE-based criteria, the test suggested in [205], based on the test proposed in [207], was chosen, i. e.:

$$T_{ij} = \max\left(0, \frac{\varepsilon\text{-RMSE}_i^2}{\varepsilon\text{-RMSE}_j^2} - F(0.05, C, C)\right), \tag{6.3}$$

with $F(0.05, C, C)$ being the F-value calculated at a 0.05 significance level and $C = 5$, the considered number of conditions (i. e. algorithms), as both degrees of freedom.[4] The hypothesis of statistical indifference is accepted if $T_{ij} = 0$, while it is rejected (i. e. ε-RMSE$_i$ performs statistically significantly worse than ε-RMSE$_j$) if $T_{ij} > 0$ [206, 207].

6.6 Results and Discussion

The results and discussion in this section is divided into two subsections: the first discussing the statistical analysis for the subjective scores alone, and the second discussing the relationships between the subjective and the objective scores.

6.6.1 Subjective-evaluation measures

The subjective scores collected with the MUSHRA were first post-screened, according to the standard [178], in order to remove those subjects (if any) assigning very high scores to the anchor, and/or very low scores to the hidden reference. One of the subjects from the native-speaker group had to be removed as he consistently rated the hidden reference lower than 90; this, possibly, indicates that the subject did not properly understand the instructions. Thus the total number of subjects used for the subsequent analysis was 20: 11 native speakers and 9 non-native speakers.

[4] The test requires the relation ε-RMSE$_i > \varepsilon$-RMSE$_j$ to hold [207].

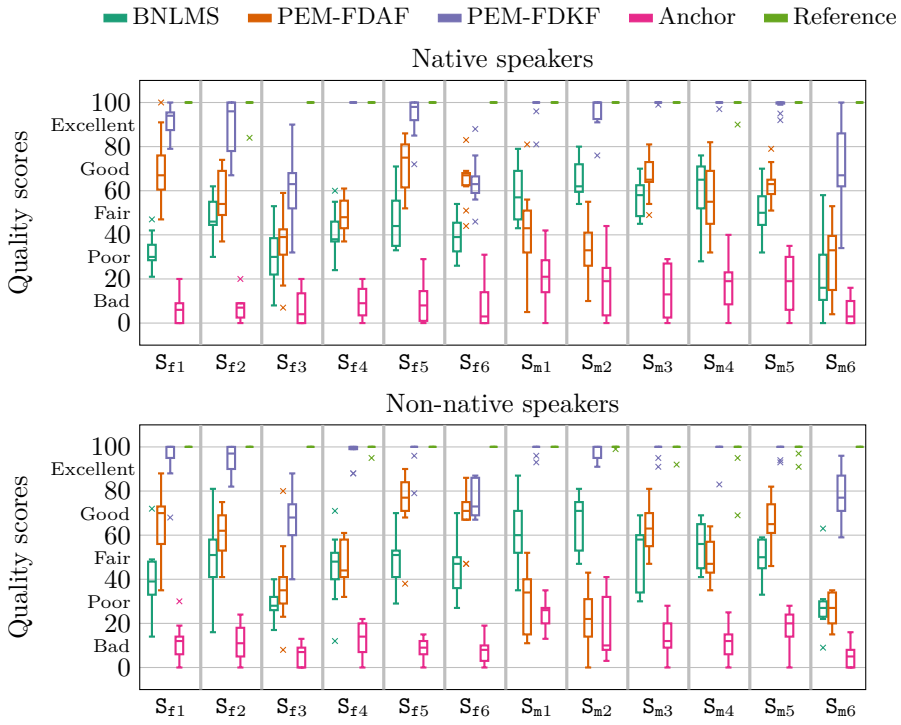


Figure 6.4: Box plots of the subjective scores for the five different conditions (i. e. algorithms) across all different speech signals. For each speech signal, the algorithms plotted from left to right reflect the ordering expressed in the legend. The results from the native speakers and the non-native speakers are kept separated.

The results for native and non-native speakers, obtained for the speech signals (i. e. $\mathtt{S_{f1}}$ to $\mathtt{S_{m6}}$), are shown by means of boxplots in fig. 6.4. The boxplots for both groups are consistent across test signals, as indicated by a monotonic increase in the scores of, in ascending order, anchor, BNLMS, PEM-FDAF, PEM-FDKF, and reference. The increase of the gain from K_2 to K_3 in $\mathtt{S_{f3}}$ to $\mathtt{S_{f4}}$ and in $\mathtt{S_{m5}}$ to $\mathtt{S_{m6}}$ can be clearly seen as all three tested algorithms show a strong reduction in the median subjective score and an increased variance. Therefore, a strong effect might be expected with gain variations at high forward path gain values. However, as mentioned in section 6.2.3, since not all the source signals were processed with K_3, no statistical analysis could be carried out using the level as an independent variable. The results for native and non-native speakers, obtained for the music signals (i. e. $\mathtt{M_1}$ to $\mathtt{M_8}$), are shown by means of boxplots

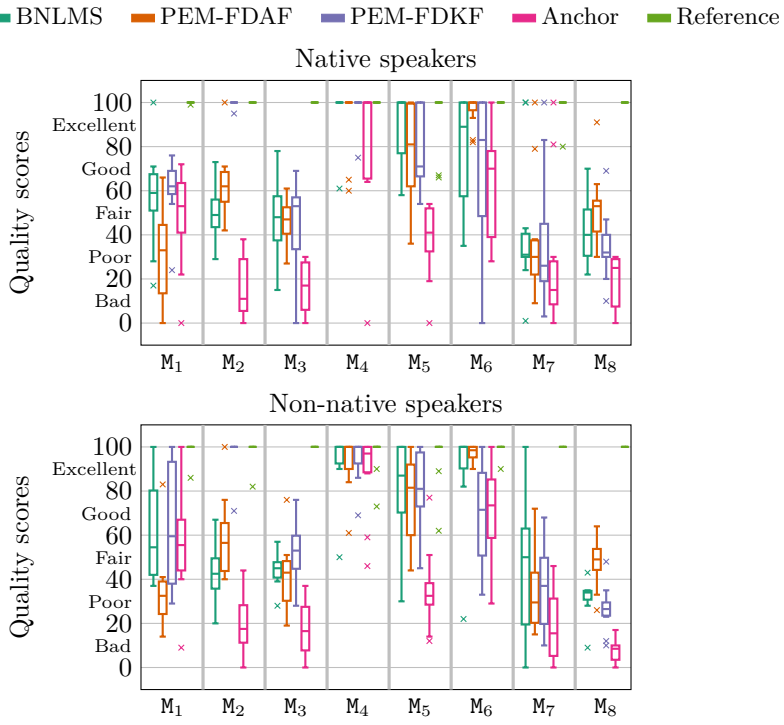


Figure 6.5: Box plots of the subjective scores for the five different conditions (i. e. algorithms) across all different music signals. For each music signal, the algorithms plotted from left to right reflect the ordering expressed in the legend. The results from the native speakers and the non-native speakers are kept separated.

in fig. 6.5. In this case, the increase observed for speech signals cannot be seen and it is noticeable how the three algorithms do not seem to consistently yield large subjective-score differences. In particular, two music signals, $\mathtt{M}_4$ and $\mathtt{M}_6$, return the highest degree of confusion between the different conditions, with more than 25 % of the subjects assigning a score ≥ 90 to the anchor. The reason of this high variability, reflecting the subjects' inability to recognize the differences between the different processing algorithms, can be explained by considering the nature of those two music signals: both are characterized by a dynamic wideband content. Therefore, the typical narrowband artifacts of an AFC scenario are likely to be masked by the simultaneous wideband frequency content. The increase of the gain from K_1 to K_2 in $\mathtt{M}_2$ to $\mathtt{M}_3$ causes an effect that is more difficult to interpret than for speech signals. The subjective scores show

Table 6.2: Results of the statistical analyses by means of the mixed-design ANOVAs. The first two columns indicate the analysis number and the effect variable The following four columns report the results of the mixed-design ANOVA analyses, i. e. the degrees of freedom, the F-values, the effect sizes η_G^2, and the corresponding Huynh-Feldt corrected p-values p_A. The three significance codes applied to the p-values are the following: $p < .001$ (**boldface**), $.001 < p < .05$ (underlined), $.05 < p < .1$ (squiggly underlined), and $p > .1$ (standard typeface).

		mixed-design ANOVA			
Analysis	Effect	(DF_n, DF_d)	F	η_G^2	p_A
1 (Speech)	NA	$(1, 18)$	4.66×10^{-4}	6.08×10^{-6}	9.83×10^{-1}
	AL	$(2, 36)$	7.33×10^{2}	7.21×10^{-1}	$\mathbf{6.89\times10^{-30}}$
	SI	$(11, 198)$	3.27×10^{1}	4.06×10^{-1}	$\mathbf{4.56\times10^{-24}}$
	NA:AL	$(2, 36)$	2.42	8.46×10^{-3}	1.03×10^{-1}
	NA:SI	$(11, 198)$	6.95×10^{-1}	1.43×10^{-2}	6.67×10^{-1}
	AL:SI	$(22, 396)$	3.19×10^{1}	3.66×10^{-1}	$\mathbf{4.85\times10^{-49}}$
	NA:AL:SI	$(22, 396)$	8.26×10^{-1}	1.47×10^{-2}	6.43×10^{-1}
2 (Speech w/o S_{f1}, S_{f4}, S_{m1}, S_{m2}, and S_{m6})	NA	$(1, 18)$	5.54×10^{-4}	7.53×10^{-6}	9.81×10^{-1}
	AL	$(2, 36)$	4.79×10^{2}	6.69×10^{-1}	$\mathbf{1.14\times10^{-26}}$
	SI	$(6, 108)$	3.15×10^{1}	4.16×10^{-1}	$\mathbf{3.05\times10^{-13}}$
	NA:AL	$(2, 36)$	7.43×10^{-1}	3.13×10^{-3}	4.83×10^{-1}
	NA:SI	$(6, 108)$	6.14×10^{-1}	1.37×10^{-2}	6.29×10^{-1}
	AL:SI	$(12, 216)$	1.49×10^{1}	1.84×10^{-1}	$\mathbf{2.24\times10^{-19}}$
	NA:AL:SI	$(12, 216)$	2.27×10^{-1}	3.42×10^{-3}	9.94×10^{-1}
3 (Music)	NA	$(1, 18)$	1.69×10^{-3}	2.78×10^{-5}	9.68×10^{-1}
	AL	$(2, 36)$	4.58	1.11×10^{-2}	1.69×10^{-2}
	SI	$(7, 126)$	5.95×10^{1}	5.58×10^{-1}	$\mathbf{1.84\times10^{-30}}$
	NA:AL	$(2, 36)$	4.04×10^{-1}	9.85×10^{-4}	6.71×10^{-1}
	NA:SI	$(7, 126)$	6.32×10^{-1}	1.32×10^{-2}	6.96×10^{-1}
	AL:SI	$(14, 252)$	1.98×10^{1}	2.34×10^{-1}	$\mathbf{2.68\times10^{-24}}$
	NA:AL:SI	$(14, 252)$	5.77×10^{-1}	8.84×10^{-3}	8.31×10^{-1}
4 (Music w/o M_4 and M_6)	NA	$(1, 18)$	3.73×10^{-2}	6.91×10^{-4}	8.49×10^{-1}
	AL	$(2, 36)$	1.79×10^{1}	5.23×10^{-2}	$\mathbf{4.57\times10^{-6}}$
	SI	$(5, 90)$	5.87×10^{1}	5.68×10^{-1}	$\mathbf{1.87\times10^{-19}}$
	NA:AL	$(2, 36)$	1.42×10^{-1}	4.38×10^{-4}	8.65×10^{-1}
	NA:SI	$(5, 90)$	6.76×10^{-1}	1.49×10^{-2}	5.92×10^{-1}
	AL:SI	$(10, 180)$	3.04×10^{1}	2.59×10^{-1}	$\mathbf{1.96\times10^{-24}}$
	NA:AL:SI	$(10, 180)$	5.79×10^{-1}	6.63×10^{-3}	7.74×10^{-1}

a very limited increase for the BNLMS, a small decrease for the PEM-FDAF, and a very strong decrease (including a large increase in the variance) for the PEM-FDKF.

The results of the statistical analysis are reported in table 6.2 in terms of signal type, effect, and results from the mixed-design ANOVA (i. e. degrees of freedom, F-values, effect sizes η_G^2, and p-values p_A), and in table 6.3 in terms of signal

Table 6.3: Results of the statistical analyses by means of the Welch-James's tests with approximate degrees of freedom. The first two columns indicate the analysis number and the effect variable The following three columns report the results of the Welch-James's tests with approximate degrees of freedom, i. e. the degrees of freedom, the values of the Welch-James's statistics T_{WJ}, and the corresponding p-values p°_{WJ}. The three significance codes applied to the p-values are the following: $p < .001$ (**boldface**), $.001 < p < .05$ (underlined), $.05 < p < .1$ (squiggly underlined), and $p > .1$ (standard typeface).

		Welch-James's test		
Analysis	Effect	(DF_n, DF_d)	T_{WJ}	p°_{WJ}
1 (Speech)	NA	$(1, 273)$	4.63×10^{-1}	4.86×10^{-1}
	AL	$(2, 182)$	1.28×10^{3}	$\mathbf{< 1\times10^{-6}}$
	SI	$(11, 109)$	4.00×10^{1}	$\mathbf{< 1\times10^{-6}}$
	NA:AL	$(2, 182)$	3.02	5.21×10^{-2}
	NA:SI	$(11, 109)$	1.22	2.38×10^{-1}
	AL:SI	$(22, 107)$	1.53×10^{1}	$\mathbf{< 1\times10^{-6}}$
	NA:AL:SI	$(22, 107)$	5.25×10^{-1}	9.4×10^{-1}
2 (Speech w/o S_{f1}, S_{f4}, S_{m1}, S_{m2}, and S_{m6})	NA	$(1, 158)$	5.68×10^{-1}	4.63×10^{-1}
	AL	$(2, 107)$	5.90×10^{2}	$\mathbf{< 1\times10^{-6}}$
	SI	$(6, 70)$	4.87×10^{1}	$\mathbf{< 1\times10^{-6}}$
	NA:AL	$(2, 107)$	1.22	3.13×10^{-1}
	NA:SI	$(6, 70)$	1.51	1.83×10^{-1}
	AL:SI	$(12, 65)$	8.92	$\mathbf{< 1\times10^{-6}}$
	NA:AL:SI	$(12, 65)$	4.48×10^{-1}	9.35×10^{-1}
3 (Music)	NA	$(1, 125)$	2.11×10^{-2}	8.86×10^{-1}
	AL	$(2, 87)$	3.83	1.8×10^{-2}
	SI	$(7, 74)$	2.86×10^{2}	$\mathbf{< 1\times10^{-6}}$
	NA:AL	$(2, 87)$	1.93×10^{-1}	8.43×10^{-1}
	NA:SI	$(7, 74)$	6.49×10^{-1}	7.11×10^{-1}
	AL:SI	$(14, 69)$	2.65×10^{1}	$\mathbf{< 1\times10^{-6}}$
	NA:AL:SI	$(14, 69)$	3.35×10^{-1}	9.8×10^{-1}
4 (Music w/o M_4 and M_6)	NA	$(1, 84)$	7.99×10^{-2}	7.68×10^{-1}
	AL	$(2, 57)$	1.42×10^{1}	$\mathbf{< 1\times10^{-6}}$
	SI	$(5, 58)$	4.02×10^{2}	$\mathbf{< 1\times10^{-6}}$
	NA:AL	$(2, 57)$	2.44×10^{-1}	7.58×10^{-1}
	NA:SI	$(5, 58)$	7.93×10^{-1}	5.17×10^{-1}
	AL:SI	$(10, 53)$	3.56×10^{1}	$\mathbf{< 1\times10^{-6}}$
	NA:AL:SI	$(10, 53)$	2.95×10^{-1}	9.81×10^{-1}

type, effect, and results of the Welch-James's tests with approximate degrees of freedom (i. e. degrees of freedom, values of the Welch-James's statistics T_{WJ}, and p-values p°_{WJ}). The analysis for the speech signals was done twice: once including all the speech signals, and once excluding those speech signals that required a post-processing in either of the baseline conditions, i. e. S_{f1}, S_{f4}, S_{m1}, S_{m2}, and S_{m6}, from now referred to as postprocessed speech signals. The analysis for the music signals was done twice: once including all the signals from M_1 to

$\mathtt{M}_8$, and once excluding $\mathtt{M}_4$ and $\mathtt{M}_6$.

Speech signals

The first mixed-design ANOVA was carried out using all the speech signals as `AU` within-subject factor, cf. top portion of table 6.2. The analysis illustrates that the scores by native and non-native speakers were not significantly different. A significant effect with $p_A < .001$ was found for `AL`, `AU` and the interaction effect between `AL` and `AU` (`AL:AU`). According to the guidelines found in literature [184], the effect size for `AL`, `AU` and `AL:AU`, can be considered as large. For instance, in the case of `AL`, $\eta^2_G = 7.21 \times 10^{-1}$ means that, considering the data from a given algorithm, we can explain 72 % of the subjective scores variability for that given algorithm evaluation. Therefore, we can confidently claim that different algorithms lead to different results. The significant effect on `AU` was to be expected, due to the signal-dependent nature of feedback and AFC artifacts. No significant effect was found for the two remaining two-way interactions, `NA:AL` and `NA:AU`, and for the single three-way interaction `NA:AL:AU`, and all the corresponding effect sizes were found to be small. The results of the Welch-James's test, cf. top portion of table 6.3, in terms of p-values significance, confirm the results of the mixed-design ANOVA for all the effects but the `NA:AL`, cf. top portions of tables 6.2 and 6.3. Given that the `AL` effect was found to be significant for both analyses, a Tukey's HSD test and a post-hoc Welch-James's test were applied to assess the statistical difference between the three tested algorithms. The mean value of the scores for the three algorithms were as follows: 46.4 for the BNLMS, 51.9 for the PEM-FDAF, and 89.5 for the PEM-FDKF. The p-values p_T obtained from the Tukey's HSD test, the p-values p^{p}_{WJ} obtained from the post-hoc Welch-James's test, and the effect sizes d were as follows: [$p_T < .001$, $p_{WJ} < .001$, $d = 0.29$] between the BNLMS and the PEM-FDAF, [$p_T < .001$, $p_{WJ} < .001$, $d = 2.56$] between the BNLMS and the PEM-FDKF, and [$p_T < .001$, $p_{WJ} < .001$, $d = 2.03$] between the PEM-FDAF and the PEM-FDKF.

The second mixed-design ANOVA was carried out using only seven of the twelve speech signals (i. e. excluding the postprocessed speech signals) as `AU` within-subject factor, cf. second portion from the top of table 6.2. The analysis illustrates that the scores by native and non-native speakers were not significantly different. A significant effect with $p_A < .001$ was found for `AL`, `AU` and `AL:AU`. The effect size for `AL` and `AU` is large, while it is small for `AL:AU`, as in the previous case where all the speech signals were included, we can confidently claim that different algorithms lead to different results. No significant effect was found for the three remaining multi-way interactions, and all the corresponding effect sizes were found to be small. The results of the

Welch-James's test, cf. second portion from the top of table 6.3, in terms of p-values significance, confirm the results of the mixed-design ANOVA, cf. second portions from the top of tables 6.2 and 6.3. The mean value of the scores for the three algorithms were as follows: 47.0 for the BNLMS, 59.1 for the PEM-FDAF, and 87.6 for the PEM-FDKF. The results of the two post-hoc tests were as follows: [$p_T < .001$, $p_{WJ} < .001$, $d = 0.75$] between the BNLMS and the PEM-FDAF, [$p_T < .001$, $p_{WJ} < .001$, $d = 2.43$] between the BNLMS and the PEM-FDKF, and [$p_T < .001$, $p_{WJ} < .001$, $d = 1.61$] between the PEM-FDAF and the PEM-FDKF.

Overall, both analyses with the speech signals suggest that there is a strong statistically significant effect for `AL`, `AU` and `AL:AU`, backed up by large effect sizes. According to the post-hoc tests results, the PEM-FDKF outperforms the other two algorithms in terms of subjectively-evaluated sound quality, given the very low p-values and the large effect sizes. As mentioned in section 6.2.4, the inclusion of the postprocessed speech signals might lead to a more conservative subjective-scores difference between the PEM-FDKF and the BNLMS/PEM-FDAF. Therefore, it is expected that an even bigger difference would be found, had the postprocessed speech signals not been included. The improvement of the PEM-FDAF over the BNLMS is significant in both analyses, but the effect size varies from small to large once the postprocessed speech signals are removed. Therefore, even though there seems to be an existing difference, strong evidence cannot be drawn from our experiment. An experiment with increased number of speech signals, or the inclusion of more speech signals processed with a lower forward path gain might strengthen such evidence.

Music signals

The third mixed-design ANOVA was carried out using the music signals as `AU` within-subject factor, cf. third portion of table 6.2. The analysis illustrates that the scores by native and non-native speakers were not significantly different. A significant effect was found for `AL` ($p_A < .05$) `AU` ($p_A < .001$) and the interaction effect `AL:AU` ($p_A < .001$). In this case, the effect size for both `AU` and `AL:AU` is large, while the effect size for `AL` is small. Compared to the analysis for the speech signals, the higher level of the p-value for `AL` ($p_A < .05$ versus $p_A < .001$) is likely caused by the higher variability of the subjective scores and the similarity of the performance of the different algorithms. Nevertheless, the null hypothesis (i. e. equal mean for the three algorithms) could still be rejected using a significance threshold of 5 %. However, the very small effect size suggests that the overall score difference caused by the three tested algorithms might be, in reality, negligible. No significant effect was found for the three remaining multi-way interactions, and all the corresponding effect sizes were found to be

small. The results of the Welch-James's test, cf. third portion of table 6.3, in terms of p-values significance, confirm the results of the mixed-design ANOVA, cf. third portions of tables 6.2 and 6.3. The mean value of the scores for the three algorithms were as follows: 62.3 for the BNLMS, 61.7 for the PEM-FDAF, and 65.9 for the PEM-FDKF. The results of the two post-hoc tests were as follows: [$p_T = 9.13 \times 10^{-1}$, $p_{WJ} = 3.63 \times 10^{-1}$, $d = 0.02$] between the BNLMS and the PEM-FDAF, [$p_T = 3 \times 10^{-2}$, $p_{WJ} = 8.17 \times 10^{-2}$, $d = 0.13$] between the BNLMS and the PEM-FDKF, and [$p_T = 9.09 \times 10^{-3}$, $p_{WJ} = 6.86 \times 10^{-3}$, $d = 0.15$] between the PEM-FDAF and the PEM-FDKF.

The fourth mixed-design ANOVA was carried out using only six of the eight music signals (i. e. without $\mathtt{M}_4$ and $\mathtt{M}_6$) as `AU` within-subject factor, cf. bottom portion of table 6.2. The analysis illustrates that the scores by native and non-native speakers were not significantly different. A significant effect with $p_A < .001$ was found for `AL`, `AU`, and the interaction effect `AL:AU`. The effect size for both `AU` and `AL:AU` is still large, while effect size for `AL` is still small. Compared to the previous case, the removal of $\mathtt{M}_4$ and $\mathtt{M}_6$ leads to a lower p-value for `AL` ($p_A < .001$ versus $p_A < .05$), but the effect size remains small. This, again, suggests that the overall score difference caused by the three tested algorithms might be, in reality, negligible. No significant effect was found for the three remaining multi-way interactions, and all the corresponding effect sizes were found to be small. The results of the Welch-James's test, cf. bottom portion of table 6.3, in terms of p-values significance, confirm the results of the mixed-design ANOVA, cf. bottom portions of tables 6.2 and 6.3. The mean value of the scores for the three algorithms were as follows: 55.3 for the BNLMS, 53.1 for the PEM-FDAF, and 62.5 for the PEM-FDKF. The results of the two post-hoc tests were as follows: [$p_T = 3.44 \times 10^{-1}$, $p_{WJ} = 1.73 \times 10^{-1}$, $d = 0.08$] between the BNLMS and the PEM-FDAF, [$p_T < .001$, $p_{WJ} < .001$, $d = 0.27$] between the BNLMS and the PEM-FDKF, and [$p_T < .001$, $p_{WJ} < .001$, $d = 0.35$] between the PEM-FDAF and the PEM-FDKF.

Overall, both analyses with the music signals suggest that there is a strong statistically significant effect for `AU` and `AL:AU`, backed up by large effect sizes. For `AL`, the p-values are $< .001$ only when removing $\mathtt{M}_4$ and $\mathtt{M}_6$, but the effect sizes remain small. The post-hoc tests suggest that there might be a slight perceptual sound-quality improvement when using the PEM-FDKF over the PEM-FDAF and the BNLMS, but the effect is minimal. This suggests that the statistically significant improvement obtained for the PEM-FDKF over the other two algorithms in terms of subjectively-evaluated sound quality might not be relevant. The removal of $\mathtt{M}_4$ and $\mathtt{M}_6$ does decrease the p-values but does not, substantially, change the effect size. On the other hand, there is no significant difference between the perceived sound quality when using either the PEM-FDAF or the BNLMS. Unlike for speech signals, no postprocessing was

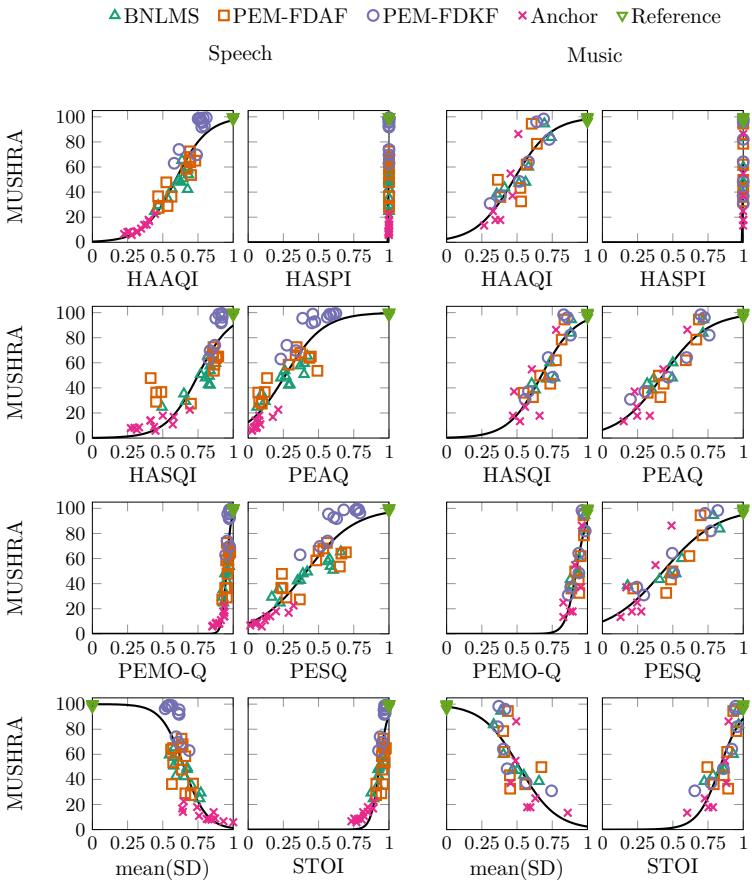


Figure 6.6: Scatter plots of subjective scores versus objective scores, including the sigmoid mapping calculated for the set of data within each scatter plot.

necessary for music signals.

6.6.2 Objective measures and performance criteria

The results of the comparisons between the different objective scores and the subjective scores are shown in the scatter plots in fig. 6.6. Each of the subplots depicts the scores for a single objective measure plotted against the subjective scores. Overlayed to each scatter plot is the sigmoid function fitted on the data.

The scores obtained using either speech or music signals are kept separated (first two and last two columns from the left, respectively).

The results for speech signals show that, for most of the objective measures, there is a good visual correspondence between objective and subjective scores. For three of the objective measures, i. e. HASPI, PEMO-Q, and STOI, it can be observed how most of the objective scores are located in the right-hand side of the plot, suggesting that the predictions yield values in a narrower range. This effect alone is not enough to dismiss the predictive capabilities, but it makes it harder to visually evaluate how good the prediction is. The remaining objective measures yield scores distributed along the whole range. It can also be observed how for some of the objecive measures, e. g. HAAQI, HASQI, PEAQ, PESQ, and mean(SD), the results from single algorithms are lumped on a side of the sigmoid fitting (see, e. g., the PEM-FDKF results for HAAQI).

The results for music signals show a similar outcome: a good visual correspondence between objective and subjective scores in most cases, except for HASPI and PEMO-Q. STOI, in this case, yields scores more widely distributed than for the speech signals.

In order to quantify these aspects, the results in terms of the different performance criteria, defined in section 6.5, are shown in table 6.4. The best performances, separately calculated for speech and music signals, are reported in bold (up to statistical indifference) for two performance criteria, as done in [205], i. e. r_{sig} and ε-RMSE. The cases that yielded a value of ε-RMSE $= 0$ are indicated by '-'.

The results for speech signals, in terms of correlation-based criteria, show that HASPI performs the worst in terms of r and r_{sig}. The mean(SD), too, performs poorly compared to the other objective measures, all yielding values of $|r| > 0.8$ and values of $r_{\text{sig}} \geq 0.9$. Nevertheless, the monotonicity is maintained by all objective measures, all yielding $|r_{\text{spear}}| \geq 0.9$. The best-performing objective measure, using r_{sig} as a predictor, is the HAAQI reaching $r_{\text{sig}} = 0.96$, statistically indifferent from the values of PEAQ, PEMO-Q, and PESQ. In terms of RMSE-based criteria, the situation is unchanged, with those objective measures yielding high correlations also yielding low RMSE and ε-RMSE. In this case, too, the HAAQI yields the best performance with a ε-RMSE $= 0.06$, statistically different from all the other ε-RMSE values. Although HAAQI and PEAQ were mainly designed to predict subjective scores for music signals, they seem to yield very good predictive capabilities for speech signals. These results suggest that HAAQI, PEAQ, PEMO-Q, or PESQ might be a better choice over SD to objectively evaluate the speech quality when using AFC algorithms. Unexpectedly, STOI, too, performed well even though it has primarily been used to predict speech intelligibility.

Table 6.4: Results obtained using the different performance criteria when comparing objective and subjective scores, separately calculated for speech and music signals. With the **boldface** we indicate the best performances (statistically indifferent) computed for r_{sig} and ε-RMSE.

Signal type	Objective measure	$\lvert r \rvert$	$\lvert r_{\text{spear}} \rvert$	r_{sig}	RMSE	ε-RMSE
speech	HAAQI	0.95	0.96	**0.96**	7.60	**0.06**
	HASPI	0.56	0.95	0.58	34.58	25.61
	HASQI	0.90	0.94	0.92	11.85	4.44
	PEAQ	0.88	0.94	**0.94**	7.39	-
	PEMO-Q	0.89	0.94	**0.93**	9.81	1.23
	PESQ	0.94	0.95	**0.95**	7.45	0.39
	mean(SD)	0.78	0.90	0.88	12.08	3.35
	STOI	0.88	0.95	0.92	14.15	3.64
music	HAAQI	0.88	0.91	0.91	2.89	-
	HASPI	0.74	0.91	0.73	23.85	16.36
	HASQI	0.92	0.93	**0.92**	5.02	-
	PEAQ	0.94	0.95	**0.96**	2.99	-
	PEMO-Q	0.89	0.90	0.89	7.53	**2.78**
	PESQ	0.93	0.93	**0.93**	5.43	-
	mean(SD)	0.81	0.88	0.85	3.98	-
	STOI	0.88	0.90	0.88	7.15	**3.16**

The results for music signals, in terms of correlation-based criteria, show, too, that HASPI performs the worst for all three criteria, followed by the mean(SD). The best-performing objective measure is PEAQ reaching $r_{\text{sig}} = 0.96$, statistically indifferent from the values of HASQI and PESQ. In terms of RMSE-based criteria, the RMSE values are, in this case, lower than those found for speech signals. This can be understood by considering the aforementioned phenomenon involving the outliers lumped on one side of the sigmoid fittings. These outliers are more common in speech signals than in music signals, and the RMSE is very sensitive to outliers, leading to lower RMSE values compared to those found for speech signals. When comparing the RMSE values with the correlation-based measures, we see again a correspondence in the objective measures yielding high correlations and low RMSE and ε-RMSE. However, the ε-RMSE is less informative in this case as for most of the objective measures the variability is too large to yield a useful ε-RMSE. PEMO-Q and STOI yields the best, statistically indifferent, performance in terms of ε-RMSE.

Overall, these results point out that several existing objective measures can effectively be used to predict the perceptually-evaluated sound quality of signals

subject to acoustic feedback and processed by AFC algorithms. Although some objective measures perform better than others, in terms of a given performance criteria, the differences might be very limited and not univocally represented by the results from all the other performance criteria. Therefore, a single objective measure cannot be selected as the best performing one, in terms of all the different criteria and for both speech and music signals. Some of the objective measures tested in this study are widespread, e. g. PEAQ, PESQ, and STOI, and their use could simplify the assessment of AFC algorithms working in conjunction with other processing stages usually included in a HA, e. g. noise reduction and compression algorithms. Nevertheless, further tests including more subjects and test signals are necessary to successfully generalize the outcomes of this study.

6.7 Conclusions

In this study, we have presented an experiment to assess the perceptually-evaluated sound quality of three different AFC algorithms processing speech and music signals.

The analysis of the collected data have shown that both native and non-native speakers perform in a statistically indifferent way for both music and speech signals. The analysis on speech signals and on a partial set of the music signals has shown that both `AU` and `AL` cause a highly significant effect, but the effect size for `AL` is large only for speech signals. The effect on `AL` has subsequently been studied with a post hoc analysis, in order to find a ranking of the algorithms performance: the PEM-FDKF has been found to be the algorithm providing best performance in both cases, where the large effect size for speech signals suggests the generalizability of the result.

The subjective scores have been compared to objective scores obtained with a set of well-known objective measures. A very high correlation, using different correlation coefficients, has been found when using the objective measures designed to assess perceptual sound quality. Low RMSE-based values have been found for those objective measures yielding high correlations. These results suggest that different existing objective measures might be used to predict the sound-quality degradation caused by acoustic feedback and AFC artifacts, for both speech and music signals, potentially reducing the testing times and costs of assessing AFC algorithms.

Chapter 7

Conclusion

This thesis presented research originating from a collection of multidisciplinary objectives related to the topic of signal processing, in terms of feedback cancellation, for implantable hearing devices. The different major objectives were the following: to characterize and analyse the feedback path in a novel implantable hearing device, the Codacs direct acoustic cochlear implant (DACI), to develop improved adaptive feedback cancellation (AFC) algorithms for hearing devices, and to subjectively assess the impact of one of the developed AFC algorithms, the prediction-error method (PEM)-based frequency-domain Kalman filter (PEM-FDKF), on sound quality.

The order of presentation reflects, somehow, the order in which the task of designing a novel AFC algorithm for (novel) hearing devices should be tackled, starting from the assessment of whether or not, and if so when and how, the feedback problem manifests itself, moving then to the actual AFC algorithm design, and concluding with the subjective assessment of the AFC algorithm performance in terms of sound quality. Naturally, the final step in a real application scenario would be the actual implementation and validation of the AFC algorithm, which were not treated in this thesis but that will likely be pursued as future work.

Measurements and analysis of feedback for the Codacs DACI

The first major objective of this work was that of measuring the feedback signals of a novel implantable hearing device, the Codacs DACI, in order to obtain a characterization of the corresponding feedback path.

Chapter 2 described the measurements and analysis of feedback and nonlinearities for the Codacs DACI. The feedback measurements provided an answer to each of the four initial research questions. Namely, the feedback path was found to be specimen-dependent, the effects of an incorrect implantation were found to strongly affect the measured feedback, the response of the actuator was found to be nonlinear at the measured levels, and a difference between the structure-borne mechanical (i. e. tissular) feedback component and the acoustic (i. e. airborne) feedback component was noted. It was postulated that the feedback path variations caused by an incorrect implantation, might have, potentially, been related to the feedback artifacts mentioned in the literature. Additionally, it was suggested how the characterization procedure performed during the cadaver head measurements could be further tested to discriminate between correct and incorrect implantations. Finally, it was proposed that the feedback path recorded on a given specimen could be exploited to facilitate the design of a dedicated AFC algorithm.

Chapter 3 cannot be shared publicly due to confidentiality obligations.

Development of improved AFC algorithms for hearing devices

The second major objective of this work was that of designing novel AFC algorithms, in order to improve the performance and reduce the computational complexity of existing algorithms. In chapters 4 and 5, the design of two such novel AFC algorithms was discussed.

Chapter 4 presented the design, the complexity analysis, and the evaluation by means of computer simulations, of the frequency-domain prediction-error-method-based adaptive filter (FD-PEMAF), a variant of the PEM-based frequency-domain adaptive filter (PEM-FDAF) relying on frequency-domain PEM-based decorrelation. The algorithm was shown to have comparable performance with the PEM-FDAF, in terms of estimation error, achievable amplification, and sound-quality degradation, while allowing for a reduced computational complexity. Thus, such an algorithm could be a good candidate for use in low-power devices, such as hearing devices. An unconstrained version of the algorithm was also presented, resulting in strongly degraded performance albeit allowing for an even further reduction in the computational complexity. Therefore the latter simplification was discouraged.

Chapter 5 presented two other AFC algorithms, specifically tailored for hearing devices. Both the proposed algorithms, i. e. the PEM-FDKF and the PEM-based partitioned-block frequency-domain Kalman filter (PEM-PBFDKF), combine good decorrelation properties, by means of the PEM-based signal prewhitening, fast convergence, good tracking behavior, and low computational

complexity by means of the frequency-domain Kalman filter (FDKF). The proposed algorithms were tested in different scenarios, considering both stationarity and non-stationarity of the feedback path, given the relevance of feedback path variability in a hearing device real-life scenario. Furthermore, the PEM-PBFDKF was shown to yield low algorithmic delay by exploiting a partitioned block (PB) implementation. Finally, a theoretical derivation of the identifiability conditions, necessary to apply the PEM-FDKF in a closed-loop scenario, was provided.

Both the FD-PEMAF and the two Kalman-filter-based algorithms could be eligible candidates for a hearing device implementation. Although the FD-PEMAF might represent a more suitable solution in the short term, due to its low computational complexity, the Kalman-filter-based algorithms represent a better alternative to current algorithms in the long term. Therefore, as soon as the complexity of the PEM-FDKF, or of the PEM-PBFDKF, becomes acceptable for a hearing device implementation, it could represent a viable alternative to existing AFC algorithms.

Subjective assessment of AFC algorithms for hearing devices

The third major objective of this work was that of providing a validation of the improved performance of newly derived PEM-FDKF, in terms of perceptual sound quality. In order to do this, a subjective listening test was designed and executed in order to evaluate the perceptual sound quality of audio signals affected by feedback- and AFC-related acoustic artifacts.

In chapter 6, a perceptual sound-quality assessment of different AFC algorithms was presented. The assessment was carried out by means of listening tests involving human test subjects, and it suggested that the PEM-FDKF performs significantly better, in terms of sound-quality preservation, compared to the PEM-FDAF and the block normalized least mean squares (BNLMS). The improved performance is large for speech signals and smaller, yet still statistically significant, for music signals. These findings validate the simulation results presented in chapter 5.

Additionally, a comparison between the subjective perceptual sound-quality scores and the results obtained with a batch of existing objective scores showed that a good subjective-to-objective correlation is obtained. This observation might motivate the use of such objective scores to assess, in a quick and inexpensive way, the performance of novel AFC algorithms.

Suggestions for future research

As mentioned at the beginning of this chapter, the contributions presented in this thesis align with the research process carried out when developing a novel AFC algorithm for a (novel) hearing device. The lack of the final algorithm validation through the implementation in the hearing device points to a few straightforward future research directions.

For instance, for the AFC algorithms derived in this thesis to be successfully implemented into the two novel implantable hearing devices considered in this thesis, some additional steps should be made:

- A complexity analysis based on the specific hardware constraints of the hearing device should be carried out. This is needed because the real-time version presented in the thesis was only validated on a high-performance real-time machine without the low-power and the (possibly) lower-precision constraint of a hearing device. In this way, it would be possible to verify whether or not the algorithm can be successfully run in real time on the specific hearing device and, if that were not to be the case, some improved real-time implementation strategies or algorithmic simplifications could be considered.
- A version of the algorithm using fixed-point arithmetic, based on the specific architecture of the hearing device sound processor, should be implemented, if necessary. This is because, to date, the sound processors of many hearing devices still work with a fixed-point arithmetic. The performance of such a fixed-point implementation should, in turn, be validated by a comparison with the performance obtained in a simulation scenario.
- A test of the possible conflicts between the existing algorithms running in the hearing device sound processor and the proposed AFC algorithms should be carried out. This is important, given that other adaptive algorithms might be running simultaneously when the hearing device operates, most notably adaptive noise reduction algorithms.

Besides the discussed steps, several additional points related to each of the three parts of the thesis offer, too, interesting ideas for further research directions.

Feedback measurements and analysis

Considering the aforementioned points, a working real-time version of the proposed AFC algorithms could be implemented on the Codacs DACI and be

used for additional tests on cadaver heads. By first keeping the AFC algorithm inactive, a validation of the estimated feedback margins could be carried out, followed by a validation of its performance, upon the AFC algorithm activation. The frequency-domain feedback paths collected in the measurements described in chapter 2 could be used as prior knowledge and be incorporated into the algorithms in order to speed up convergence. In case the specimen-dependent feedback variability is too large compared to the AFC algorithm flexibility, a quick estimation procedure using the exponential sine sweep (ESS) technique on the considered cadaver head would provide a better prior knowledge and, likely, help to improve the algorithmic performance.

Once all these tests have been carried out, a similar procedure involving feedback estimation, AFC algorithm tuning, and validation could be carried out in implanted patients.

A more complicated approach could involve feedback measurements of the cadaver heads in different conditions in order to simulate changes in the feedback path, similarly to what has been done with mannequin heads [119]. This would be done, e. g., by placing a highly acoustically reflective surface close to the behind-the-ear (BTE) sound processor. In such a scenario, it would be interesting to investigate how the different conditions impact the feedback measurement and, if possible, how they impact the different feedback components.

Developments of the AFC algorithms

The two novel AFC algorithms presented in chapters 4 and 5 are both based on the PEM. Therefore, for the sake of simplicity, only innovations retaining this specific stage of the proposed AFC algorithms will be discussed.

One of the main assumptions made in both algorithms is the possibility of modeling the source signal using a time-varying autoregressive (AR) model. This is a simplistic, yet functional, model when applied to speech; however, for music the model is, in most cases, an oversimplification. Therefore, a first way to improve the proposed algorithms could be that of introducing a cascade of different source signal models, able to deal with tonal audio signals. Such models have already been studied in an AFC context [109], although using a time-domain implementation. Therefore, it might be easier to apply such models to the PEM-FDKF, which relies on time-domain PEM-based signal prewhitening, than to the FD-PEMAF, which relies on frequency-domain PEM-based signal prewhitening. The inclusion of more complex source signal models will, inevitably, lead to an increased computational complexity; perhaps too large for a current state-of-the-art hearing device. Nevertheless, it could still

be used in other applications, such as in feedback reduction for public address (PA) systems or for future generations of hearing devices.

Another modeling aspect that could be further studied is the simple time-varying description attributed to the feedback path in the PEM-FDKF. Using a simple first-order Markov model as an abstraction of the true feedback path dynamics is an oversimplification when considering abrupt changes that might happen in a real-life scenario for a hearing-device patient, e. g. wearing a hat, or being close to a highly acoustically reflective surface. It is, hence, sensible to think about a model improvement in order to take care of both smooth and abrupt feedback path variations.

Finally, the proposed AFC algorithms could be used in conjunction with other feedback control algorithms, in order to exploit the strengths of different methods to tackle the same problem. For instance, an AFC algorithm and a notch-filter-based feedback suppression algorithm (as well as another kind of postfilter) could be used, so that in those situations where the AFC algorithm alone is not able to stop the howling, the notch-filter-based feedback suppression algorithm could intervene.[1]

Subjective assessment of AFC algorithms

One of the problems encountered when designing the listening test discussed in chapter 6, was the need to introduce a post processing stage and manually reduce the gain of the loudspeaker signals in those portions where howling was occurring. Given that such a task was signal-dependent and algorithm-dependent, it was quite laborious. An improvement over the manual method, would be that of using a howling detector to locate the portion of the signals exhibiting howling and, based on this information, apply the postprocessing automatically.[2] Notwithstanding, the inclusion of a postprocessing stage inherently introduces a bias on sound-quality subjective ratings. This is due to the fact that signals that would be rated with a lower score because of the perceptually annoying howling artifact are rated with a higher score after the howling effect is reduced by the postprocessing. However, lowering the forward path gain until all the tested AFC algorithms do not exhibit any instability, would not result in a fair comparison, as the performance at high forward path gain is an important aspect that was sought for. Therefore, an interesting research direction could

[1] The use of a postfilter to improve the performance of an AFC algorithm is currently being investigated in a master's thesis project entitled "Perceptually motivated postfilter design in adaptive feedback cancellation algorithms".

[2] New machine-learning-based techniques are currently being investigated to improve the performance of automatic howling detectors in a master's thesis project entitled "Machine learning for notch-filter-based howling suppression and acoustic feedback control".

involve the design of a better sound-quality evaluation procedure capable of taking all these aspects into account.

Additionally, given that this was a pilot study in the sound-quality evaluation of AFC algorithms, more data should be collected. Besides the benefit of validating the subjective findings of the research presented chapter 6, testing more subjects and more conditions (i. e. audio signals and algorithms) would also allow to validate the correlations found between subjective and objective measurements. As a means for more easily collecting data from new subjects, a crowdsourced perceptual sound-quality evaluation test might be a viable opportunity.[3]

Finally, given that the AFC algorithms described in this thesis are mainly tailored for hearing-device applications, listening tests with hearing-device users should be organized. However, this requires, especially for implantable hearing-device users, a real-time implementation of the AFC algorithms embedded in the hearing-device sound processor that must be, in first place, tested and validated to comply with safety standards.

[3]The crowdsourcing of a speech intelligibility test is currently being investigated in a master's thesis project entitled "Development, implementation, and evaluation of a web-based speech intelligibility test".

Bibliography

[1] World Health Organization, "Deafness and hearing loss." `http://www.who.int/mediacentre/factsheets/fs300/en/`. Date last accessed: Jun. 30, 2017.

[2] D. Moore, P. Fuchs, A. Rees, A. Palmer, and C. Plack, *Oxford Handbook of Auditory Science: Hearing*. Oxford, United Kingdom: OUP Oxford, 2010.

[3] N. Verhaert, C. Desloovere, and J. Wouters, "Acoustic hearing implants for mixed hearing loss: a systematic review," *Otol. Neurotol.*, vol. 34, pp. 1201–1209, Sep. 2013.

[4] H. Bernhard, C. Stieger, and Y. Perriard, "Design of a semi-implantable hearing device for direct acoustic cochlear stimulation," *IEEE Trans. Biomed. Eng.*, vol. 58, pp. 420–428, Feb. 2011.

[5] H. Dillon, *Hearing Aids*. Turramurra, Australia: Boomerang Press, 2001.

[6] S. Doclo, W. Kellermann, S. Makino, and S. E. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," *IEEE Signal Process. Mag.*, vol. 32, pp. 18–30, Mar. 2015.

[7] J. J. Shea Jr, "Forty years of stapes surgery," *Otol. Neurotol.*, vol. 19, pp. 52–55, Jan. 1998.

[8] R. A. Goldenberg and J. R. Emmet, "Current use of implants in middle ear surgery," *Otol. Neurotol.*, vol. 22, pp. 145–152, Mar. 2001.

[9] H. A. Jenkins, J. K. Niparko, W. H. Slattery, J. G. Neely, and J. M. Fredrickson, "Otologics middle ear transducer™ ossicular stimulator: Performance results with varying degrees of sensorineural hearing loss," *Acta Otolaryngol.*, vol. 124, pp. 391–394, Sep. 2004.

[10] L. Bruschini, S. Berrettini, F. Forli, A. Murri, and D. Cuda, "The Carina© middle ear implant: surgical and functional outcomes," *Eur. Arch. Otorhinolaryngol.*, vol. 273, pp. 3631–3640, Nov. 2016.

[11] B. Fraysse, J.-P. Lavieille, S. Schmerber, V. Enée, E. Truy, C. Vincent, F. M. Vaneecloo, and O. Sterkers, "A multicenter study of the vibrant soundbridge middle ear implant: early clinical results and experience," *Otol. Neurotol.*, vol. 22, pp. 952–961, Nov. 2001.

[12] C. M. Luetje, D. Brackman, T. J. Balkany, J. Maw, R. S. Baker, D. Kelsall, D. Backous, R. Miyamoto, S. Parisier, and A. Arts, "Phase III clinical trial results with the vibrant soundbridge implantable middle ear hearing device: A prospective controlled multicenter study," *Otolaryngol. Head Neck Surg.*, vol. 126, pp. 97–107, Feb. 2002.

[13] K. Kroll, I. L. Grant, and E. Javel, "The Envoy® totally implantable hearing system, St. Croix Medical," *Trends Amplif.*, vol. 6, pp. 73–80, Jun. 2002.

[14] D. A. Chen, D. D. Backous, M. A. Arriaga, R. Garvin, D. Kobylek, T. Littman, S. Walgren, and D. Lura, "Phase 1 clinical trial results of the Envoy system: A totally implantable middle ear device for sensorineural hearing loss," *Otolaryngol. Head Neck. Surg.*, vol. 131, pp. 904–916, Dec. 2004.

[15] J. Hough, P. Matthews, M. W. Wood, and R. K. Dyer Jr, "Middle ear electromagnetic semi-implantable hearing device: results of the phase II SOUNDTEC direct system clinical trial," *Otol. Neurotol.*, vol. 23, pp. 895–903, Nov. 2002.

[16] S. Pelosi, M. L. Carlson, and M. E. Glasscock, "The Ototronix MAXUM system: Indications, techniques, and outcomes," *Oper. Tech. Otolaryngol.–Head Neck Surg.*, vol. 25, pp. 338–343, Dec. 2014.

[17] J. B. Hunter, M. L. Carlson, and M. E. Glasscock, "The Ototronix MAXUM middle ear implant for severe high-frequency sensorineural hearing loss: Preliminary results," *Laryngoscope*, vol. 126, pp. 2124–2127, Mar. 2016.

[18] C. N. Kahue, M. L. Carlson, J. A. Daugherty, D. S. Haynes, and M. E. Glasscock III, "Middle ear implants for rehabilitation of sensorineural hearing loss: a systematic review of FDA approved devices," *Otol. Neurotol.*, vol. 35, pp. 1228–1237, Aug. 2014.

[19] A. Tjellström and B. Håkansson, "The bone-anchored hearing aid. Design principles, indications, and long-term clinical results," *Otolaryngol. Clin. North Am.*, vol. 28, pp. 53–72, Feb. 1995.

[20] A. Tjellström, B. Håkansson, and G. Granström, "Bone-anchored hearing aids: Current status in adults and children," *Otolaryngol. Clin. North Am.*, vol. 2, pp. 337–364, Apr. 2001.

[21] S. Reinfeldt, B. Håkansson, H. Taghavi, and M. Eeg-Olofsson, "New developments in bone-conduction hearing implants: a review," *Medical Devices: Evidence and Research*, vol. 8, pp. 79–93, Jan. 2015.

[22] M. C. Flynn, A. Hedin, G. Halvarsson, T. Good, and A. Sadeghi, "Hearing performance benefits of a programmable power BAHA® sound processor with a directional microphone for patients with a mixed hearing loss," *Clin. Exp. Otorhinolaryngol.*, vol. 5, pp. S76–S81, Apr. 2012.

[23] P. Westerkull, "The Ponto bone-anchored hearing system," in *Implantable Bone Conduction Hearing Aids*, vol. 71, pp. 32–40, Karger Publishers, 2011.

[24] K. Kozlowski and D. R. Friedland, "Implantable hearing devices," *Curr. Surg. Rep.*, vol. 2, Jul. 2014.

[25] M. Barbara, M. Perotti, B. Gioia, L. Volpini, and S. Monini, "Transcutaneous bone-conduction hearing device: audiological and surgical aspects in a first series of patients with mixed hearing loss," *Acta Otolaryngol.*, vol. 133, pp. 1058–1064, Jun. 2013.

[26] A. Huber, J. Sim, Y. Xie, M. Chatzimichalis, O. Ullrich, and C. Röösli, "The bonebridge: Preclinical evaluation of a new transcutaneously-activated bone anchored hearing device," *Hear. Res.*, vol. 301, pp. 93–99, Jul. 2013.

[27] G. Sprinzl, T. Lenarz, A. Ernst, R. Hagen, A. Wolf-Magele, H. Mojallal, I. Todt, R. Mlynski, and M. D. Wolframm, "First European multicenter results with a new transcutaneous bone conduction hearing implant system: short-term safety and efficacy," *Otol. Neurotol.*, vol. 34, pp. 1076–1083, Aug. 2013.

[28] B. Håkansson, S. Reinfeldt, M. Eeg-Olofsson, P. Östli, H. Taghavi, J. Adler, J. Gabrielsson, S. Stenfelt, and G. Granström, "A novel bone conduction implant (BCI): Engineering aspects and pre-clinical studies," *Int. J. Audiol.*, vol. 49, pp. 203–215, Jan. 2010.

[29] H. Taghavi, B. Håkansson, S. Reinfeldt, M. Eeg-Olofsson, and S. Akhshijan, "Feedback analysis in percutaneous bone-conduction device and bone-conduction implant on a dry cranium," *Otol. Neurotol.*, vol. 33, pp. 413–420, Apr. 2012.

[30] M. Eeg-Olofsson, B. Håkansson, S. Reinfeldt, H. Taghavi, H. Lund, K.-J. Fredén Jansson, E. Håkansson, and J. Stalfors, "The bone conduction implant—First implantation, surgical and audiologic aspects," *Otol. Neurotol.*, vol. 35, pp. 679–685, Apr. 2014.

[31] S. Reinfeldt, B. Håkansson, H. Taghavi, K.-J. Fredén Jansson, and M. Eeg-Olofsson, "The bone conduction implant: Clinical results of the first six patients," *Int. J. Audiol.*, vol. 54, pp. 408–416, Jan. 2015.

[32] H. Taghavi, B. Håkansson, S. Reinfeldt, M. Eeg-Olofsson, K.-J. Fredén Jansson, E. Håkansson, and B. Nasri, "Technical design of a new bone conduction implant (BCI) system," *Int. J. Audiol.*, vol. 54, pp. 736–744, Jun. 2015.

[33] R. Häusler, C. Stieger, H. Bernhard, and M. Kompis, "A novel implantable hearing system with direct acoustic cochlear stimulation," *Audiol. Neurotol.*, vol. 13, pp. 247–256, Nov. 2008.

[34] R. Häusler, S. Colburn, and E. Marr, "Sound localization in subjects with impaired hearing," *Acta Otolaryngol.*, vol. 96, pp. 1–62, Feb. 1983.

[35] C. C. Dunn, R. S. Tyler, and S. A. Witt, "Benefit of wearing a hearing aid on the unimplanted ear in adult users of a cochlear implant," *J. Speech Lang. Hear. Res.*, vol. 48, pp. 668–680, Jun. 2005.

[36] V. Koci, J. Seebacher, V. Weichbold, P. Zorowka, A. Wolf-Magele, G. Sprinzl, and K. Stephan, "Improvement of sound source localization abilities in patients bilaterally supplied with active middle ear implants," *Acta Otolaryngol.*, vol. 136, pp. 692–698, Jul. 2016.

[37] T. van den Bogaert, T. J. Klasen, M. Moonen, L. Van Deun, and J. Wouters, "Horizontal localization with bilateral hearing aids: without is better than with," *J. Acoust. Soc. Amer.*, vol. 119, pp. 515–526, Jan. 2006.

[38] T. Ching, E. van Wanrooy, and H. Dillon, "Binaural-bimodal fitting or bilateral implantation for managing severe to profound deafness: A review," *Trends Amplif.*, vol. 11, pp. 161–192, Sep. 2007.

[39] M. F. Dorman, S. Cook, A. Spahr, T. Zhang, L. Loiselle, D. Schramm, J. Whittingham, and R. Gifford, "Factors constraining the benefit to speech understanding of combining information from low-frequency hearing and a cochlear implant," *Hear. Res.*, vol. 322, pp. 107–111, Apr. 2015.

[40] R. English, K. Plant, M. Maciejczyk, and R. Cowan, "Fitting recommendations and clinical benefit associated with use of the NAL-NL2 hearing-aid prescription in Nucleus cochlear implant recipients," *Int. J. Audiol.*, vol. 55, pp. S45–S50, Feb. 2016.

[41] B. C. Moore, "Binaural sharing of audio signals: Prospective benefits and limitations," *Hear. J.*, vol. 60, pp. 46–48, Nov. 2007.

[42] J. M. Kates, *Signal Processing for Hearing Aids*, pp. 235–277. Boston, MA: Springer, 2002.

[43] V. Hamacher, J. Chalupper, J. Eggers, E. Fischer, U. Kornagel, H. Puder, and U. Rass, "Signal processing in high-end hearing aids: State of the art, challenges, and future trends," *EURASIP J. Appl. Signal Process.*, vol. 2005, pp. 2915–2929, Jan. 2005.

[44] V. Hamacher, "Comparison of advanced monaural and binaural noise reduction algorithms for hearing aids," in *Proc. 2002 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, vol. 4, pp. 4008–4011, May 2002.

[45] T. J. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *IEEE Trans. Signal Process.*, vol. 55, pp. 1579–1585, Apr. 2007.

[46] T. van Waterschoot and M. Moonen, "Fifty years of acoustic feedback control: State of the art and future challenges," *Proc. IEEE*, vol. 99, pp. 288–327, Feb. 2011.

[47] M. C. Killion, A. Van Halteren, S. Stenfelt, and D. M. Warren, "Hearing aid transducers," in *Hearing Aids* (G. R. Popelka, B. C. J. Moore, R. R. Fay, and A. N. Popper, eds.), pp. 59–92, Cham, Switzerland: Springer, 2016.

[48] T. Lenarz, J. W. Zwartenkot, C. Stieger, B. Schwab, E. A. Mylanus, M. Caversaccio, M. Kompis, A. F. Snik, C. D'hondt, and H. Mojallal, "Multicenter study with a direct acoustic cochlear implant," *Otol. Neurotol.*, vol. 34, pp. 1215–1225, Sep. 2013.

[49] J. Agnew, "Acoustic feedback and other audible artifacts in hearing aids," *Trends Amplif.*, vol. 1, pp. 45–82, Jun. 1996.

[50] A. F. Snik, E. A. Mylanus, and C. W. Cremers, "The bone-anchored hearing aid compared with conventional hearing aids: audiologic results and the patients' opinions," *Otolaryngol. Clin. North Am.*, vol. 28, pp. 78–83, Feb. 1995.

[51] K. Chung, "Challenges and recent developments in hearing aids: Part II. Feedback and occlusion effect reduction strategies, laser shell manufacturing

processes, and other signal processing technologies," *Trends Amplif.*, vol. 8, pp. 125–164, Fall 2004.

[52] L. L. N. Wong, L. Hickson, and B. McPherson, "Hearing aid satisfaction: What does research from the past 20 years say?," *Trends Amplif.*, vol. 7, pp. 117–161, Fall 2003.

[53] S. Kochkin, "MarkeTrak VII: Customer satisfaction with hearing instruments in the digital age," *Hear. J.*, vol. 58, pp. 30–43, Sep. 2005.

[54] S. Kochkin, "MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly increasing," *Hear. J.*, vol. 63, pp. 19–32, Jan. 2010.

[55] J. Hellgren, T. Lunner, and S. Arlinger, "Variations in the feedback of hearing aids," *J. Acoust. Soc. Amer.*, vol. 106, pp. 2821–2833, Nov. 1999.

[56] J. M. Kates, "Room reverberation effects in hearing aid feedback cancellation," *J. Acoust. Soc. Amer.*, vol. 109, pp. 367–378, Jan. 2001.

[57] B. Rafaely, M. Roccasalva-Firenze, and E. Payne, "Feedback path variability modeling for robust hearing aids," *J. Acoust. Soc. Amer.*, vol. 107, pp. 2665–2673, May 2000.

[58] S. Kampe and M. Wynne, "The influence of venting on the occlusion effect," *Hear. J.*, vol. 49, pp. 59–68, 1996.

[59] L. Revit, "Two techniques for dealing with the occlusion effect," *Hear. Instruments*, vol. 43, pp. 16–16, 1992.

[60] A. Winkler, M. Latzel, and I. Holube, "Open versus closed hearing-aid fittings: A literature review of both fitting approaches," *Trends Hear.*, vol. 20, pp. 1–13, Jan. 2016.

[61] J. M. Kates, "Feedback cancellation in hearing aids: results from a computer simulation," *IEEE Trans. Signal Process.*, vol. 39, pp. 553–562, Mar. 1991.

[62] A. J. Bosman, A. F. Snik, E. A. Mylanus, and C. W. Cremers, "Fitting range of the BAHA Intenso," *Int. J. Audiol.*, vol. 48, pp. 346–352, Jun. 2009.

[63] M. J. Park, J. R. Lee, C. J. Yang, M. H. Yoo, I. S. Jin, C. H. Choi, and H. J. Park, "Amplification of transcutaneous and percutaneous bone-conduction devices with a test-band in an induced model of conductive hearing loss," *Int. J. Audiol.*, vol. 55, pp. 653–657, Jun. 2016.

[64] M. J. Syms and K. E. Hernandez, "Bone conduction hearing," *Otolaryngol. Head Neck. Surg.*, vol. 150, pp. 866–871, Mar. 2014.

[65] F. Memari, A. Asghari, A. Daneshi, and A. Jalali, "Safety and patient selection of totally implantable hearing aid surgery: Envoy system, Esteem," *Eur. Arch. Otorhinolaryngol.*, vol. 268, pp. 1421–1425, Oct. 2011.

[66] M. Lachowska, K. Niemczyk, A. Yazbeck, K. Morawski, and A. Bruzgielewicz, "First bilateral simultaneous implantation with fully implantable middle ear hearing device," *Arch. Med. Sci.*, vol. 8, pp. 736–742, Apr. 2012.

[67] M. Debeaupte, E. Decullier, S. Tringali, A. Devèze, T. Mom, V. Darrouzet, and E. Truy, "Evolution of the reliability of the fully implantable middle ear transducer over successive generations," *Otol. Neurotol.*, vol. 36, pp. 625–630, Apr. 2015.

[68] G. Bernardi, T. van Waterschoot, M. Moonen, J. Wouters, M. Hillbratt, and N. Verhaert, "Measurement and analysis of feedback and nonlinearities for the Codacs direct acoustic cochlear implant," *IEEE Access*, vol. 5, pp. 8702–8713, May 2017.

[69] F. V. Hunt, "Acoustic vs acoustical," *J. Acoust. Soc. Am.*, vol. 27, pp. 975–976, Sep. 1955.

[70] T. van Waterschoot, G. Rombouts, and M. Moonen, "Optimally regularized adaptive filtering algorithms for room acoustic signal enhancement," *Signal Process.*, vol. 88, pp. 594–611, Mar. 2008.

[71] H. Schepker and S. Doclo, "Least-squares estimation of the common pole-zero filter of acoustic feedback paths in hearing aids," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 24, pp. 1334–1347, Aug. 2016.

[72] G. Bernardi, T. van Waterschoot, J. Wouters, and M. Moonen, "Adaptive feedback cancellation using a partitioned-block frequency-domain Kalman filter approach with PEM-based signal prewhitening," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, pp. 1480–1494, Sep. 2017.

[73] M. R. Schroeder, "Improvement of feedback stability of public address systems by frequency shifting," *J. Audio Eng. Soc.*, vol. 10, pp. 108–109, Oct. 1962.

[74] M. Schroeder, "Improvement of acoustic-feedback stability by frequency shifting," *J. Acoust. Soc. Amer.*, vol. 36, pp. 1718–1724, Sep. 1964.

[75] D. K. Bustamante, T. L. Worrall, and M. J. Williamson, "Measurement and adaptive suppression of acoustic feedback in hearing aids," in *Proc. 1989 IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, pp. 2017–2020, May 1989.

[76] J. Maxwell and P. Zurek, "Reducing acoustic feedback in hearing aids," *IEEE Trans. Speech Audio Process.*, vol. 3, pp. 304–313, Jul. 1995.

[77] S. Armstrong, F. Sykes, D. Brown, and J. Ryan, "Digital hearing aid system." Patent, US6937738 B2, Aug. 2005.

[78] P. Gil-cacho, T. van Waterschoot, M. Moonen, and S. H. Jensen, "Regularized Adaptive Notch Filters for Acoustic Howling Suppression," in *Proc. 17th Eur. Signal Process. Conf. (EUSIPCO)*, (Glasgow, Scotland), pp. 2574–2578, Sep. 2009.

[79] P. Estermann and A. Kaelin, "Feedback cancellation in hearing aids: results from using frequency-domain adaptive filters," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 2, pp. 257–260, May 1994.

[80] J. Kates, "Feedback cancellation apparatus and methods." Patent, US6072884 A, Jun. 2000.

[81] B. Rafaely and M. Roccasalva-Firenze, "Control of feedback in hearing aids—A robust filter design approach," *IEEE Trans. Speech Audio Process.*, vol. 8, pp. 754–756, Nov. 2000.

[82] J. Hellgren, "Analysis of feedback cancellation in hearing aids with filtered-x LMS and the direct method of closed loop identification," *IEEE Trans. Speech Audio Process.*, vol. 10, pp. 119–131, Feb. 2002.

[83] A. Spriet, I. Proudler, M. Moonen, and J. Wouters, "Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal," *IEEE Trans. Signal Process.*, vol. 53, pp. 3749–3763, Oct. 2005.

[84] A. Spriet, G. Rombouts, M. Moonen, and J. Wouters, "Adaptive feedback cancellation in hearing aids," *J. Franklin Instit.*, vol. 343, pp. 545–573, Aug. 2006.

[85] G. Ma, F. Gran, F. Jacobsen, and F. Agerkvist, "Adaptive feedback cancellation with band-limited lpc vocoder in digital hearing aids," *IEEE Trans. Audio Speech Lang. Process.*, vol. 19, pp. 677–687, May 2011.

[86] K. Ngo, T. van Waterschoot, M. G. Christensen, M. Moonen, and S. H. Jensen, "Improved prediction error filters for adaptive feedback cancellation in hearing aids," *Signal Process.*, vol. 93, pp. 3062–3075, Nov. 2013.

[87] F. Strasser and H. Puder, "Adaptive feedback cancellation for realistic hearing aid applications," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, pp. 2322–2333, Dec 2015.

[88] H. Schepker, L. T. T. Tran, S. Nordholm, and S. Doclo, "Improving adaptive feedback cancellation in hearing aids using an affine combination of filters," in *Proc. 2016 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 231–235, Mar 2016.

[89] J. E. Greenberg, P. M. Zurek, and M. Brantley, "Evaluation of feedback-reduction algorithms for hearing aids," *J. Acoust. Soc. Amer.*, vol. 108, pp. 2366–2376, Nov. 2000.

[90] S. Haykin, *Adaptive filter theory.* Englewood Cliffs, NJ: Prentice Hall, 2002.

[91] M. G. Siqueira, R. Speece, E. Petsalis, A. Alwan, S. Soli, and S. Gao, "Subband adaptive filtering applied to acoustic feedback reduction in hearing aids," in *Conf. Rec. 30th Asilomar Conf. Signals, Syst. Computers*, vol. 1, pp. 788–792, Nov. 1996.

[92] M. G. Siqueira and A. Alwan, "Steady-state analysis of continuous adaptation in acoustic feedback reduction systems for hearing-aids," *IEEE Trans. Speech Audio Process.*, vol. 8, pp. 443–453, Jul. 2000.

[93] G. Rombouts, T. van Waterschoot, K. Struyve, and M. Moonen, "Acoustic feedback cancellation for long acoustic paths using a nonstationary source model," *IEEE Trans. on Signal Process.*, vol. 54, pp. 3426–3434, Sep. 2006.

[94] A. Spriet, S. Doclo, M. Moonen, and J. Wouters, "Feedback control in hearing aids," in *Springer Handbook of Speech Processing* (J. Benesty, M. M. Sondhi, and Y. A. Huang, eds.), pp. 979–1000, Berlin, Germany: Springer, 2008.

[95] M. Rotaru, F. Albu, and H. Coanda, "A variable step size modified decorrelated NLMS algorithm for adaptive feedback cancellation in hearing aids," in *Proc. 10th Int. Symp. Electron. Telecommun. (ISETC)*, pp. 263–266, Nov. 2012.

[96] D. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech Audio Process.*, vol. 8, pp. 508–518, Sep. 2000.

[97] J. Hellgren and U. Forssell, "Bias of feedback cancellation algorithms in hearing aids based on direct closed loop identification," *IEEE Trans. Speech Audio Process.*, vol. 9, pp. 906–913, Nov. 2001.

[98] J. Benesty and D. Morgan, "Frequency-domain adaptive filtering revisited, generalization to the multi-channel case, and application to acoustic echo cancellation," in *Proc. 2000 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, vol. 2, pp. 789–792, 2000.

[99] T. van Waterschoot, G. Rombouts, P. Verhoeve, and M. Moonen, "Double-talk-robust prediction error identification algorithms for acoustic echo cancellation," *IEEE Trans. Signal Process.*, vol. 55, pp. 846–858, Mar. 2007.

[100] J. Cioffi and T. Kailath, "Fast, recursive-least-squares transversal filters for adaptive filtering," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 32, pp. 304–337, Apr. 1984.

[101] D. T. M. Slock and T. Kailath, "Numerically stable fast transversal filters for recursive least squares adaptive filtering," *IEEE Trans. Signal Process.*, vol. 39, pp. 92–114, Jan. 1991.

[102] M. Guo, S. Jensen, and J. Jensen, "Novel acoustic feedback cancellation approaches in hearing aid applications using probe noise and probe noise enhancement," *IEEE Trans. Audio Speech Lang. Process.*, vol. 20, pp. 2549–2563, Nov. 2012.

[103] C. R. C. Nakagawa, S. Nordholm, and W. Y. Yan, "Feedback cancellation with probe shaping compensation," *IEEE Signal Process. Lett.*, vol. 21, pp. 365–369, Mar. 2014.

[104] G. Schmidt and T. Haulick, "Signal processing for in-car communication systems," *Signal Process.*, vol. 86, pp. 1307–1326, Jun. 2006.

[105] M. Guo, S. H. Jensen, J. Jensen, and S. L. Grant, "On the use of a phase modulation method for decorrelation in acoustic feedback cancellation," in *Proc. 20th Eur. Signal Process. Conf. (EUSIPCO)*, pp. 2000–2004, Aug. 2012.

[106] F. J. van der Meulen, S. Kamerling, and C. P. Janse, "A new way of acoustic feedback suppression," in *Preprints AES 104th Conv.*, (Amsterdam, The Netherlands), May 1998.

[107] L. Ljung, ed., *System Identification: Theory for the User (2nd Edition)*. Upper Saddle River, NJ: Prentice Hall, 1999.

[108] U. Forssell and L. Ljung, "Closed-loop identification revisited," *Automatica*, vol. 35, pp. 1215–1241, Jul. 1999.

[109] T. van Waterschoot and M. Moonen, "Adaptive feedback cancellation for audio applications," *Signal Process.*, vol. 89, pp. 2185–2201, Nov. 2009.

[110] J. M. Gil-Cacho, T. van Waterschoot, M. Moonen, and S. H. Jensen, "Wiener variable step size and gradient spectral variance smoothing for double-talk-robust acoustic echo cancellation and acoustic feedback cancellation," *Signal Process.*, vol. 104, pp. 1–14, Nov. 2014.

[111] E. R. Ferrara, "Fast implementations of LMS adaptive filters," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 28, pp. 474–475, Aug. 1980.

[112] G. Clark, S. Mitra, and S. Parker, "Block implementation of adaptive digital filters," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 29, pp. 744–752, Jul. 1981.

[113] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Process. Mag.*, vol. 9, pp. 14–37, Jan. 1992.

[114] P. Sommen, P. van Gerwen, H. Kotmans, and A. Janssen, "Convergence analysis of a frequency-domain adaptive filter with exponential power averaging and generalized window function," *IEEE Trans. Circuits Syst.*, vol. 34, pp. 788–798, Jul. 1987.

[115] J.-S. Soo and K. Pang, "Multidelay block frequency domain adaptive filter," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, pp. 373–376, Feb. 1990.

[116] H. Buchner, J. Benesty, and W. Kellermann, "Generalized multichannel frequency-domain adaptive filtering: efficient realization and application to hands-free speech communication," *Signal Processing*, vol. 85, no. 3, pp. 549–570, 2005.

[117] M. Guo, T. Elmedyb, S. Jensen, and J. Jensen, "Analysis of acoustic feedback/echo cancellation in multiple-microphone and single-loudspeaker systems using a power transfer function method," *IEEE Trans. Signal Process.*, vol. 59, pp. 5774–5788, Dec. 2011.

[118] A. Spriet, M. Moonen, and J. Wouters, "Evaluation of feedback reduction techniques in hearing aids based on physical performance measures," *J. Acoust. Soc. Amer.*, vol. 128, pp. 1245–1261, Sep. 2010.

[119] N. Madhu, J. Wouters, A. Spriet, T. Bisitz, V. Hohmann, and M. Moonen, "Study on the applicability of instrumental measures for black-box evaluation of static feedback control in hearing aids," *J. Acoust. Soc. Amer.*, vol. 130, pp. 933–947, Aug 2011.

[120] T. Bisitz, N. Madhu, A. Spriet, J. Wouters, M. Moonen, B. Kollmeier, and V. Hohmann, "Relation between different measures of feedback in hearing aids and gain measures derived from percentile analysis," *Acta Acust. united Ac.*, vol. 98, pp. 676–684, Jul. 2012.

[121] A. Spriet, K. Eneman, M. Moonen, and J. Wouters, "Objective measures for real-time evaluation of adaptive feedback cancellation algorithms in hearing aids," in *Proc. 16th Eur. Signal Process. Conf. (EUSIPCO)*, (Lausanne, Switzerland), Aug. 2008.

[122] J. M. Kates and K. H. Arehart, "The Hearing-Aid Speech Quality Index (HASQI) version 2," *J. Audio Eng. Soc*, vol. 62, pp. 99–117, Feb. 2014.

[123] A. J. Manders, D. M. Simpson, and S. L. Bell, "Objective prediction of the sound quality of music processed by an adaptive feedback canceller," *IEEE Trans. Audio Speech Lang. Process.*, vol. 20, pp. 1734–1745, Aug. 2012.

[124] G. Grimm, V. Hohmann, and B. Kollmeier, "Increase and subjective evaluation of feedback stability in hearing aids by a binaural coherence-based noise reduction scheme," *IEEE Trans. Audio Speech Lang. Process.*, vol. 17, pp. 1408–1419, Sep. 2009.

[125] E. E. Johnson, T. A. Ricketts, and B. W. Hornsby, "The effect of digital phase cancellation feedback reduction systems on amplified sound quality," *J. Am. Acad. Audiol.*, vol. 18, pp. 404–416, May 2007.

[126] R. Vicen-Bueno, A. Martinez-Leira, R. Gil-Pita, and M. Rosa-Zurera, "Modified LMS-based feedback-reduction subsystems in digital hearing aids based on WOLA filter bank," *IEEE Trans. Instrum. Meas.*, vol. 58, pp. 3177–3190, Sep. 2009.

[127] K. Ngo, A. Spriet, M. Moonen, J. Wouters, and S. H. Jensen, "A combined multi-channel Wiener filter-based noise reduction and dynamic range compression in hearing aids," *Signal Processing*, vol. 92, pp. 417–426, Feb. 2012.

[128] A. Pandey and V. J. Mathews, "Low-delay signal processing for digital hearing aids," *IEEE Trans. Audio Speech Lang. Process.*, vol. 19, pp. 699–710, May 2011.

[129] C. R. C. Nakagawa, S. Nordholm, and W. Y. Yan, "Dual microphone solution for acoustic feedback cancellation for assistive listening," in *Proc. 2012 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 149–152, Mar. 2012.

[130] M. Guo, S. H. Jensen, and J. Jensen, "Evaluation of state-of-the-art acoustic feedback cancellation systems for hearing aids," *J. Audio Eng. Soc.*, vol. 61, pp. 125–137, Apr. 2013.

[131] Y. FanChiang, C. W. Wei, Y. L. Meng, Y. W. Lin, S. J. Jou, and T. S. Chang, "Low complexity formant estimation adaptive feedback cancellation for hearing aids using pitch based processing," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, pp. 1248–1259, Aug. 2014.

[132] S. Bech and N. Zacharov, *Perceptual Audio Evaluation: Theory, Method and Application*. Chichester, NY: Wiley, 2006.

[133] T. Lenarz, N. Verhaert, C. Desloovere, J. Desmet, C. D'hondt, J. González, E. Kludt, A. Macías, H. Skarżyński, P. Van de Heyning, C. Vyncke, and A. Wasowski, "A comparative study on speech in noise understanding with a direct acoustic cochlear implant in subjects with severe to profound mixed hearing loss," *Audiol. Neurotol.*, vol. 19, pp. 164–174, Jul. 2014.

[134] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Preprints AES 108th Conv.*, (Paris, France), Feb. 2000. AES Preprint 5093.

[135] ASTM Standard F2504-05, "Standard practice for describing system output of implantable middle ear hearing devices." ASTM International, 2014.

[136] A. Torras-Rosell and F. Jacobsen, "A new interpretation of distortion artifacts in sweep measurements," *J. Audio Eng. Soc.*, vol. 59, pp. 283–289, Jun. 2011.

[137] A. Torras-Rosell, "Methods of measuring impulse responses in architectural acoustics," Master's thesis, DTU, Dept. Electr. Eng., Lyngby, 2009.

[138] G.-B. Stan, J.-J. Embrechts, and D. Archambeau, "Comparison of different impulse response measurement techniques," *J. Audio Eng. Soc.*, vol. 50, no. 4, pp. 249–262, 2002.

[139] M. Grossöhmichen, R. Salcher, H.-H. Kreipe, T. Lenarz, and H. Maier, "The Codacs™ direct acoustic cochlear implant actuator: Exploring alternative stimulation sites and their stimulation efficiency," *PLoS ONE*, vol. 10, pp. 1–20, Mar. 2015.

[140] D. Havelock, S. Kuwano, and M. Vorländer, *Handbook of Signal Processing in Acoustics*. New York, NY: Springer, 2008.

[141] J. W. Zwartenkot, A. F. Snik, E. A. Mylanus, and J. J. Mulder, "Amplification options for patients with mixed hearing loss," *Otol. Neurotol.*, vol. 35, no. 2, pp. 221–226, 2014.

[142] Brüel & Kjær, *Microphone Handbook for the Falcon™ Range of Microphone Products*, 1995.

[143] J. M. Gil-Cacho, T. van Waterschoot, M. Moonen, and S. H. Jensen, "Transform domain prediction error method for improved acoustic echo and feedback cancellation," in *Proc. 20th Eur. Signal Process. Conf. (EUSIPCO)*, (Bucharest, Romania), pp. 2422–2426, Aug. 2012.

[144] J. M. Gil-Cacho, T. van Waterschoot, M. Moonen, and S. H. Jensen, "A frequency-domain adaptive filter (FDAF) prediction error method (PEM) framework for double-talk-robust acoustic echo cancellation," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, pp. 2074–2086, Dec. 2014.

[145] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, pp. 561–580, Apr. 1975.

[146] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, *Advances in Network and Acoustic Echo Cancellation*. Berlin, Germany: Springer, 2001.

[147] C. R. C. Nakagawa, S. Nordholm, and W. Y. Yan, "Analysis of two microphone method for feedback cancellation," *IEEE Signal Process. Lett.*, vol. 22, pp. 35–39, Jan 2015.

[148] T. van Waterschoot, G. Rombouts, and M. Moonen, "On the performance of decorrelation by prefiltering for adaptive feedback cancellation in public address systems," in *Proc. 4th IEEE Benelux Signal Process. Symp.*, (Hilvarenbeek, The Netherlands), pp. 167–170, Apr. 2004.

[149] A. Spriet, M. Moonen, and I. Proudler, "Feedback cancellation in hearing aids: An unbiased modelling approach," in *Proc. 11th Eur. Signal Process. Conf.*, (Toulouse, France), pp. 531–534, Sep. 2002.

[150] G. Bernardi, T. van Waterschoot, J. Wouters, M. Hillbratt, and M. Moonen, "A PEM-based frequency-domain Kalman filter for adaptive feedback cancellation," in *Proc. 23rd Eur. Signal Process. Conf. (EUSIPCO)*, (Nice, France), pp. 270–274, Aug. 2015.

[151] G. Enzner and P. Vary, "Frequency-domain adaptive Kalman filter for acoustic echo control in hands-free telephones," *Signal Process.*, vol. 86, pp. 1140–1156, Jun. 2006.

[152] G. Enzner, "Bayesian inference model for applications of time-varying acoustic system identification," in *Proc. 18th Eur. Signal Process. Conf. (EUSIPCO)*, (Aalborg, Denmark), pp. 2126–2130, Aug. 2010.

[153] A. Jain, S. Goel, K. Nathwani, and R. M. Hegde, "Robust acoustic echo cancellation using Kalman filter in double talk scenario," *Speech Commun.*, vol. 70, pp. 65–75, 2015.

[154] G. Bernardi, T. van Waterschoot, J. Wouters, and M. Moonen, "An all-frequency-domain adaptive filter with PEM-based decorrelation for acoustic feedback control," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA '15)*, (New Paltz, NY), pp. 1–5, Oct. 2015.

[155] E. Moulines, O. Ait Amrane, and Y. Grenier, "The generalized multidelay adaptive filter: structure and convergence analysis," *IEEE Trans. Signal Process.*, vol. 43, pp. 14–28, Jan. 1995.

[156] B. H. Nitsch, "A frequency-selective stepfactor control for an adaptive filter algorithm working in the frequency domain," *Signal Process.*, vol. 80, pp. 1733–1745, Sep 2000.

[157] G. W. Enzner, *A Model Based Optimum Filtering Approach to Acoustic Echo Control: Theory and Practice*. PhD thesis, RWTH, Institut für Nachrichtengeräte und Datenverarbeitung (IND), Aachen, Germany, 2006.

[158] M. L. Valero, E. Mabande, and E. A. Habets, "A state-space partitioned-block adaptivefilter for echo cancellation using inter-band correlations in the Kalman gain computation," in *Proc. 2015 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 599–603, April 2015.

[159] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.

[160] F. Kuech, E. Mabande, and G. Enzner, "State-space architecture of the partitioned-block-based acoustic echo controller," in *Proc. 2014 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 1295–1299, May 2014.

[161] S. Haykin, A. Sayed, J. Zeidler, P. Yee, and P. Wei, "Adaptive tracking of linear time-variant systems by extended RLS algorithms," *IEEE Trans. Signal Process.*, vol. 45, pp. 1118–1128, May 1997.

[162] H. W. Sorenson, "Least-squares estimation: from Gauss to Kalman," *IEEE Spectrum*, vol. 7, pp. 63–68, July 1970.

[163] J. B. Rawlings and D. Q. Mayne, *Model predictive control: Theory and design*. Madison, WI: Nob Hill Publishing, LLC, 2009.

[164] A. H. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering," *IEEE Signal Process. Mag.*, vol. 11, pp. 18–60, July 1994.

[165] F. Faccenda, S. Squartini, E. Principi, L. Gabrielli, and F. Piazza, "A real-time dual-channel speech reinforcement system for intra-cabin communication," *J. Audio Eng. Soc.*, vol. 61, pp. 889–910, Nov. 2013.

[166] G. Rombouts, T. van Waterschoot, and M. Moonen, "Robust and efficient implementation of the PEM-AFROW algorithm for acousic feedback cancellation," *J. Audio Eng. Soc.*, vol. 55, no. 11, pp. 955–966, 2007.

[167] P. Sommen, "On the convergence properties of a partitioned block frequency domain adaptive filter (PBFDAF)," in *Proc. 5th Eur. Signal Process. Conf.*, (Barcelona, Spain), pp. 201–204, Sep. 1990.

[168] H. Buchner, J. Benesty, T. Gansler, and W. Kellermann, "Robust extended multidelay filter and double-talk detector for acoustic echo cancellation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, pp. 1633–1644, Sep. 2006.

[169] M. A. Stone and B. C. Moore, "Tolerable hearing aid delays. I. estimation of limits imposed by the auditory path alone using simulated hearing losses," *Ear Hearing*, vol. 20, no. 3, pp. 182–191, 1999.

[170] Bang and Olufsen, "Music for archimedes." CD B&O 101, 1992.

[171] J. M. Kates and K. H. Arehart, "The Hearing-Aid Speech Quality Index (HASQI)," *J. Audio Eng. Soc.*, vol. 58, no. 5, pp. 363–381, 2010.

[172] S. Kochkin, "Customer satisfaction with hearing instruments in the digital age," *Hear. J.*, vol. 58, pp. 30–37, Sep. 2005.

[173] P. C. Loizou, *Speech Enhancement. Theory and Practice.* Taylor and Francis, 2007.

[174] ITU-T Rec. P.862, "Perceptual Evaluation of Speech Quality (PESQ): An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs." International Telecommunication Union, Feb. 2001.

[175] CEI/IEC 61672-1:2013, "Electroacoustics - Sound level meters - Part 1: Specifications." International Electrotechnical Commission, Sep. 2013.

[176] A. van Wieringen and J. Wouters, "LIST and LINT: sentences and numbers for quantifying speech understanding in severely impaired listeners for flanders and the netherlands," *Int. J. Audiol.*, vol. 47, no. 6, pp. 348–355, 2008.

[177] S. Jansen, R. Koning, J. Wouters, and A. van Wieringen, "Development and validation of the leuven intelligibility sentence test with male speaker (LIST-m)," *Int. J. Audiol.*, vol. 53, no. 1, pp. 55–59, 2014.

[178] ITU-T Rec. BS.1534-3, "Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA)." International Telecommunication Union, Oct. 2015.

[179] "Music Information Retrieval Evaluation eXchange (MIREX 2006)," 2007. Date last accessed: Mar. 01, 2017.

[180] M. R. Leek, M. R. Molis, L. R. Kubli, and J. B. Tufts, "Enjoyment of music by elderly hearing-impaired listeners," *J. Am. Acad. Audiol.*, vol. 19, no. 6, pp. 519–526, 2008.

[181] ITU-T Rec. BS.1116-3, "Method for the Subjective Assessment of Small Impairments in Audio Systems." International Telecommunication Union, Feb. 2015.

[182] S. Kraft and U. Zölzer, "BeaqleJS: HTML5 and JavaScript based framework for the subjective evaluation of audio quality," in *Linux Audio Conf.*, (Karlsruhe, Germany), Nov. 2014.

[183] T. Francart, A. Van Wieringen, and J. Wouters, "APEX 3: a multi-purpose test platform for auditory psychophysical experiments," *J. Neurosci. Methods*, vol. 172, no. 2, pp. 283–293, 2008.

[184] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Mahwah, NJ, USA: Lawrence Erlbaum Associates, 1988.

[185] G. M. Sullivan and R. Feinn, "Using effect size–or why the P value is not enough," *J. Grad. Med. Educ.*, vol. 4, no. 3, pp. 279–282, 2012.

[186] M. Schoeffler, A. Silzle, and J. Herre, "Evaluation of spatial/3D audio: Basic audio quality versus quality of experience," *IEEE J. Sel. Topics Signal Process.*, vol. 11, pp. 75–88, Feb. 2017.

[187] G. V. Glass, P. D. Peckham, and J. R. Sanders, "Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance," *Rev. Edu. Res.*, vol. 42, pp. 237–288, Sep. 1972.

[188] L. M. Lix, J. C. Keselman, and H. J. Keselman, "Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance f test," *Rev. Edu. Res.*, vol. 66, pp. 579–619, Dec. 1996.

[189] H. Finch, "Comparison of the performance of nonparametric and parametric MANOVA test statistics when assumptions are violated," *Methodology*, vol. 1, pp. 27–38, September 2005.

[190] R. Wilcox, *Introduction to Robust Estimation and Hypothesis Testing, 3rd Edition*. Waltham, MA: Academic Press, 2012.

[191] P. Mair and R. Wilcox, *WRS2: A Collection of Robust Statistical Methods*, May 2017. R package version 0.9-2.

[192] P. J. Villacorta, *welchADF: Welch-James Statistic for Robust Hypothesis Testing under Heterocedasticity and Non-Normality*, Oct. 2017. R package version 0.2.

[193] P. J. Villacorta, "The welchADF package for robust hypothesis testing in unbalanced multivariate mixed models with heteroscedastic and non-normal data," *R J.*, Oct. 2017.

[194] J. M. Kates and K. H. Arehart, "The hearing-aid audio quality index (HAAQI)," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 24, pp. 354–365, Feb. 2016.

[195] J. M. Kates and K. H. Arehart, "The Hearing-Aid Speech Perception Index (HASPI)," *Speech Communication*, vol. 65, pp. 75–93, 2014.

[196] ITU-R Rec. BS.1387-1, "Method for objective measurements of perceived audio quality." International Telecommunication Union, Nov. 2001.

[197] R. Huber and B. Kollmeier, "PEMO-Q – A new method for objective audio quality assessment using a model of auditory perception," *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, pp. 1902–1911, Nov. 2006.

[198] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ) - A new method for speech quality assessment of telephone networks and codecs," in *Proc. 2001 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, vol. 2, pp. 749–752, May 2001.

[199] C. Taal, R. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time and frequency weighted noisy speech," *IEEE Trans. Audio Speech Lang. Process.*, vol. 19, pp. 2125–2136, Sept 2011.

[200] M. Holters and U. Zölzer, "GstPEAQ - an open source implementation of the PEAQ algorithm," in *Proc. 18th Int. Conf. Digital Audio Effects (DAFx-15)*, vol. 4, (Trondheim, Norway), Nov. 2015.

[201] T. Dau, D. Püschel, and A. Kohlrausch, "A quantitative model of the "effective" signal processing in the auditory system. I. Model structure," *J. Acoust. Soc. Amer.*, vol. 99, pp. 3615–3622, Jun 1996.

[202] V. Emiya, E. Vincent, N. Harlander, and V. Hohmann, "Subjective and objective quality assessment of audio source separation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 19, pp. 2046–2057, Sep. 2011.

[203] ITU-T Rec. P.862.1, "Mapping function for transforming P.862 raw result scores to MOS-LQO." International Telecommunication Union, Nov. 2003.

[204] ITU-T Rec. P.862.2, "Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs." International Telecommunication Union, Nov. 2007.

[205] T. H. Falk, V. Parsa, J. F. Santos, K. Arehart, O. Hazrati, R. Huber, J. M. Kates, and S. Scollie, "Objective quality and intelligibility prediction for users of assistive listening devices: Advantages and limitations of existing tools," *IEEE Signal Process. Mag.*, vol. 32, pp. 114–124, Mar. 2015.

[206] ITU-T Rec. P.863, "Perceptual objective listening quality assessment." International Telecommunication Union, Sep. 2014.

[207] ITU-R SG 12, "Statistical evaluation procedure for P.OLQA v.1.0.." International Telecommunication Union, Mar. 2009.

Curriculum Vitae



Giuliano Bernardi was born in Asolo, Italy in 1987. He received the M.Sc. in Engineering Acoustics from Denmark Technical University (DTU), Denmark, in 2011, and the MEng in Bioengineering from University of Padua, Italy, in 2012.

Currently, he is a PhD student at the STADIUS research division within the Department of Electrical Engineering at KU Leuven. His research focuses on acoustic feedback control, specifically for hearing-aid applications.

Publication List

International Journal Papers

1. **G. Bernardi**, T. van Waterschoot, M. Moonen, J. Wouters, M. Hillbratt, and N. Verhaert, "Measurement and analysis of feedback and nonlinearities for the Codacs direct acoustic cochlear implant," *IEEE Access*, vol. 5, pp. 8702–8713, May 2017.

2. **G. Bernardi**, T. van Waterschoot, J. Wouters, and M. Moonen, "Adaptive feedback cancellation using a partitioned-block frequency-domain Kalman filter approach with PEM-based signal prewhitening," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, pp. 1480–1494, Sep. 2017.

3. **G. Bernardi**, T. van Waterschoot, J. Wouters, and M. Moonen, "Subjective and Objective Sound-Quality Evaluation of Adaptive Feedback Cancellation Algorithms," Submitted for publication to *IEEE/ACM Trans. Audio Speech Lang. Process.*, Nov. 2017.

International Conference Papers

1. **G. Bernardi**, T. van Waterschoot, J. Wouters, M. Hillbratt, and M. Moonen, "A PEM-based frequency-domain Kalman filter for adaptive feedback cancellation," in *Proc. 23rd Eur. Signal Process. Conf. (EUSIPCO)*, (Nice, France), pp. 270–274, Aug. 2015.

2. **G. Bernardi**, T. van Waterschoot, J. Wouters, and M. Moonen, "An all-frequency-domain adaptive filter with PEM-based decorrelation for acoustic feedback control," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA '15)*, (New Paltz, NY), pp. 1–5, Oct. 2015.

Patents

1. **G. Bernardi**, M. Hillbratt, T. van Waterschoot, A. Gozzi, "Methods, Systems, and Devices for Adaptively Filtering Audio Signals," U.S. Patent Application US20160255446 (A1), Sep. 1, 2016.

International Conference Abstracts

1. **G. Bernardi**, T. van Waterschoot, M. Moonen, J. Wouters, J.-M. Gérard, J. Walraevens, M. Hillbratt, and N. Verhaert, "Feedback path characterization for a direct acoustic cochlear implant," in *Abstracts 2014 Int. Hearing Aid Research Conf. (IHCON)*, (Lake Tahoe, CA), Aug. 2014.

2. **G. Bernardi**, T. van Waterschoot, M. Moonen, J. Wouters, J.-M. Gérard, J. Walraevens, M. Hillbratt, and N. Verhaert, "Complementary results on the feedback path characterization for the Cochlear™ Codacs™ direct acoustic cochlear implant," in *Abstracts 2015 Conference on Implantable Auditory Prostheses (CIAP)*, (Lake Tahoe, CA), Jul. 2015.



FACULTY OF ENGINEERING SCIENCE
DEPARTMENT OF ELECTRICAL ENGINEERING
Stadius Center for Dynamical Systems,
Signal Processing and Data Analytics
Kasteelpark Arenberg 10, B-3001 Leuven
http://www.esat.kuleuven.be/stadius/