

## ABSTRACT

**M**ODERN speech communications are evolving towards a new direction which involves users in a more perceptive way. That is the *immersive experience*, which may be considered as the “last-mile” problem of telecommunications.

One of the main feature of immersive communications is the *distant-talking*, i.e. the hands-free (in the broad sense) speech communications without body-worn or tethered microphones that takes place in a multisource environment where interfering signals may degrade the communication quality and the intelligibility of the desired speech source.

In order to preserve speech quality *intelligent acoustic interfaces* may be used. An intelligent acoustic interface may comprise multiple microphones and loudspeakers and its peculiarity is to model the acoustic channel in order to adapt to user requirements and to environment conditions. This is the reason why intelligent acoustic interfaces are based on adaptive filtering algorithms.

The acoustic path modelling entails a set of problems which have to be taken into account in designing an adaptive filtering algorithm. Such problems

may be basically generated by a linear or a nonlinear process and can be tackled respectively by linear or nonlinear adaptive algorithms.

In this work we consider such modelling problems and we propose novel effective adaptive algorithms that allow acoustic interfaces to be robust against any interfering signals, thus preserving the perceived quality of desired speech signals.

As regards *linear adaptive algorithms*, a class of adaptive filters based on the sparse nature of the acoustic impulse response has been recently proposed. We adopt such class of adaptive filters, named *proportionate adaptive filters*, and derive a general framework from which it is possible to derive any linear adaptive algorithm. Using such framework we also propose some efficient proportionate adaptive algorithms, expressly designed to tackle problems of a linear nature.

On the other side, in order to address problems deriving from a nonlinear process, we propose a novel filtering model which performs a nonlinear transformations by means of *functional links*. Using such nonlinear model, we propose *functional link adaptive filters* which provide an efficient solution to the modelling of a nonlinear acoustic channel.

Finally, we introduce robust filtering architectures based on *adaptive combinations of filters* that allow acoustic interfaces to more effectively adapt to environment conditions, thus providing a powerful mean to immersive speech communications.

## ACKNOWLEDGMENTS

**T**HE accomplishment of the Ph.D. degree is cause of huge gratification to me so I would like to share my contentment with all the people who help me to attain this goal.

*First and foremost, I would like to express my sincere appreciation and gratitude to my advisor Prof. Aurelio Uncini for his valuable guidance and continuous support throughout the entire project. He transmitted his passion for research to me and taught me how to strive for excellence in my professional pursuits.*

*I am also in debt to Prof. Michele Scarpiniti for being my well-wisher. I would like to thank him for his invaluable advices, besides being an exemplary workmate he is firstly a friend.*

*Special thanks to Dr. Mauro Falcone, I am grateful to him for giving me an opportunity to work with him and to confront myself with an important corporate reality, such as the “Fondazione Ugo Bordonni”.*

*My sincere thanks also go to my group member, Prof. Raffaele Parisi for the insightful discussions and help along the way.*

## ACKNOWLEDGMENTS

---

*I am also thankful with Prof. Jerónimo Arenas García who has incisively increased my professional skills during my visiting period at “Universidad Carlos III de Madrid”, thus resulting decisive for my thesis work. It was a great pleasure to work with him both from a professional and a personal point of view. I would like to thank him and Dr. Luis Azpicueta Ruiz for their precious welcome to Madrid.*

*I would like to thank my dissertation committee members: Prof. Giovanni Iacovitti, Prof. Antonello Rizzi, Prof. Marco Listanti for taking the time to serve on my committee and for their helpful suggestions.*

*A fundamental role has been played by my flatmates and friends, whose presence is decisive not only in my work but above all in my life.*

*My best thanks are to Alessia who is continuously at my side, cheering me up in hard times and fortunately sharing with me joys and gratifications. This work is dedicated to her.*

*To make this acknowledgment complete, a special thanks to my family for their unconditional love and support, which added the best essence to my work.*

# CONTENTS

<b>Abstract</b>	<b>vii</b>
<b>Acknowledgments</b>	<b>ix</b>
<b>Contents</b>	<b>xvi</b>
<b>List of Figures</b>	<b>xx</b>
<b>List of Tables</b>	<b>xxi</b>
<b>List of Acronyms</b>	<b>xxv</b>
<b>I INTRODUCTION</b>	<b>1</b>
<b>1 Introduction and Outline</b>	<b>3</b>
1.1 Motivations . . . . .	3
1.2 Scope of the work . . . . .	5
1.3 Organization . . . . .	7
1.4 Notation . . . . .	8

<b>2</b>	<b>Intelligent Acoustic Interfaces</b>	<b>11</b>
2.1	What is an intelligent acoustic interface? . . . . .	12
2.1.1	Acoustic interfaces . . . . .	12
2.1.2	The “intelligence” in interfaces . . . . .	13
2.1.3	Human-machine interaction by intelligent acoustic interfaces . . . . .	14
2.1.4	Applications using intelligent acoustic interfaces . . . . .	14
2.2	Science and technology of intelligent acoustic interfaces . . . . .	15
2.2.1	Historical background on speech communications . . . . .	15
2.2.2	Philosophical background on intelligent interfaces . . . . .	18
2.3	Intelligent acoustic interfaces for immersive communications . . . . .	20
<b>3</b>	<b>Problem Formulations in Acoustic Modelling</b>	<b>23</b>
3.1	Main characteristics of acoustic channels . . . . .	24
3.1.1	Linearity and shift-invariance . . . . .	24
3.1.2	Modelling by FIR filters . . . . .	25
3.1.3	Time-varying AIR . . . . .	25
3.1.4	Frequency selectivity . . . . .	26
3.1.5	Reverberation time . . . . .	26
3.1.6	Channel invertibility and minimum-phase . . . . .	27
3.1.7	Multichannel diversity . . . . .	28
3.1.8	Sparse acoustic impulse response . . . . .	28
3.2	Limitations and problems in acoustic path modelling . . . . .	28
3.2.1	Linear limitations . . . . .	29
3.2.2	Nonlinear limitations . . . . .	31
3.3	Acoustic echo cancellation . . . . .	32
3.4	Performance measures . . . . .	35
3.4.1	Echo return loss enhancement . . . . .	35
3.4.2	Normalized misalignment . . . . .	37

---

<b>II</b>	<b>LINEAR ADAPTIVE ALGORITHMS</b>	<b>39</b>
<b>4</b>	<b>Fundamentals of Adaptive Filtering</b>	<b>41</b>
4.1	Introduction to adaptive filters . . . . .	41
4.1.1	Classification of adaptive filters . . . . .	42
4.2	Linear optimum filtering . . . . .	45
4.3	Gradient adaptation . . . . .	48
4.3.1	The steepest descent method . . . . .	48
4.3.2	Convergence of the steepest descent algorithm . . . . .	49
4.4	Stochastic gradient adaptive algorithms . . . . .	51
4.4.1	The Least Mean Square Algorithm . . . . .	52
4.4.2	The Normalized Least Mean Square Algorithm . . . . .	53
4.4.3	The Recursive Least Squares Algorithm . . . . .	54
4.4.4	The Affine Projection Algorithm . . . . .	57
<b>5</b>	<b>Proportionate Adaptive Algorithms</b>	<b>59</b>
5.1	Introduction . . . . .	60
5.2	An alternative perspective on proportionate adaptive filters . . . . .	63
5.2.1	General properties of adaptive algorithms . . . . .	64
5.2.2	Natural gradient adaptation . . . . .	67
5.3	Derivation of proportionate algorithms . . . . .	69
5.4	Proportionate block APA . . . . .	72
5.5	Variable step size proportionate algorithms . . . . .	73
<b>6</b>	<b>Acoustic Interfaces Exploiting Sparsity Constraints: An Experimental Study</b>	<b>79</b>
6.1	AEC experimental conditions . . . . .	80
6.2	Performance advantages of proportionate filters . . . . .	82
6.2.1	Simplest scenario: exact path modelling in absence of near-end speech . . . . .	82
6.2.2	Exact modelling scenario in adverse environment . . . . .	86
6.3	Performance analysis of VSS proportionate filters . . . . .	88

6.3.1	Under-modelling the acoustic impulse response . . . . .	89
6.3.2	Robustness against double talk . . . . .	91
<b>III</b>	<b>NONLINEAR ADAPTIVE ALGORITHMS</b>	<b>95</b>
<b>7</b>	<b>Consequence of Nonlinearities on Hands-free Acoustic Applications</b>	<b>97</b>
7.1	Limitations of acoustic interfaces due to nonlinear interfering signals . . . . .	98
7.2	Nonlinearity effect on the performance of an AEC . . . . .	100
7.2.1	Nonlinearities in the echo transmission chain . . . . .	100
7.2.2	Loudspeaker identification by means of a neural network	102
7.2.3	Performance worsening in an AEC process . . . . .	106
<b>8</b>	<b>Functional Link Adaptive Filters: A New Class of Nonlinear Filters</b>	<b>109</b>
8.1	Introduction . . . . .	110
8.2	Nonlinear system identification problem . . . . .	112
8.3	Functional Link Adaptive Filters . . . . .	114
8.3.1	Functional link approach . . . . .	114
8.3.2	Nonlinear input expansion . . . . .	116
8.3.3	FLAF learning algorithm . . . . .	117
8.3.4	Advantages and drawbacks of FLAF . . . . .	118
8.4	Choice of functional expansion type . . . . .	119
8.4.1	Choosing a proper set of functional links . . . . .	119
8.4.2	Most popular functional link sets . . . . .	120
8.5	Memory and memoryless FLAF . . . . .	123
8.5.1	Memoryless functional links . . . . .	123
8.5.2	Functional links with memory . . . . .	125
8.6	Mean-square performance analysis . . . . .	126
8.6.1	Energetic approach to performance analysis . . . . .	126
8.6.2	Derivation of the energy conservation principle . . . . .	128



<b>9</b>	<b>Functional Link Adaptive Filters for NAEC</b>	<b>131</b>
9.1	FLAFs for acoustic applications . . . . .	132
9.2	The Split Functional Link Adaptive Filter . . . . .	132
9.3	Experimental results . . . . .	135
9.3.1	Performance improvement of SFLAFs . . . . .	136
9.3.2	An effective alternative to Volterra filters . . . . .	138
9.4	Conclusions . . . . .	140
 <b>IV ROBUST ADAPTIVE FILTERING ARCHITECTURES</b>		 <b>143</b>
<b>10</b>	<b>Filtering Architectures Based on Adaptive Combination of Filters</b>	<b>145</b>
10.1	Adaptive combination of filters . . . . .	146
10.2	Convex combination of adaptive filters . . . . .	149
10.3	Adaptation of mixing parameters . . . . .	152
10.3.1	LMS adaptation of a convex combination of two filters	152
10.3.2	A normalized adaptation . . . . .	153
10.4	Conclusions . . . . .	155
<b>11</b>	<b>Combined Architectures for Adaptive Beamforming</b>	<b>157</b>
11.1	Introduction . . . . .	158
11.2	Combined microphone array beamforming architecture . . . . .	161
11.3	Adaptive combinations of MISO systems . . . . .	162
11.3.1	Convex combination of adaptive filters using APA . . .	162
11.3.2	System-by-system combined architecture . . . . .	163
11.3.3	Filter-by-filter combined architecture . . . . .	163
11.4	Multi-stage microphone array beamforming . . . . .	165
11.5	Experimental results . . . . .	168
11.5.1	Convergence performance of combined architectures .	168
11.5.2	Speech enhancement evaluation of combined beamform- ers . . . . .	173

11.6 Conclusions . . . . .	176
<b>12 Collaborative Architectures for NAEC</b>	<b>177</b>
12.1 A serious problem in NAEC . . . . .	178
12.2 Collaborative FLAF . . . . .	180
12.3 Block-based collaborative FLAF . . . . .	183
12.4 Experimental results . . . . .	186
12.5 Conclusions . . . . .	189
<b>V CONCLUSIONS</b>	<b>191</b>
<b>13 Conclusions and Outlooks</b>	<b>193</b>
<b>References</b>	<b>210</b>
<b>Index</b>	<b>211</b>

## LIST OF FIGURES

2.1	An acoustic interface. . . . .	12
2.2	The <i>tin can telephone</i> , or also known as <i>lover's phone</i> , connected two diaphragms with a taut string or wire, which transmitted sound by mechanical vibrations from one to the other along the wire. . . . .	16
2.3	Immersive speech communication in multisource environment.	21
3.1	An acoustic interface. . . . .	29
3.2	Microphone-loudspeaker acoustic coupling. . . . .	33
3.3	Processing scheme of an acoustic echo canceller. . . . .	34
3.4	Limitation effects on the achievable ERLE. . . . .	36
4.1	Scheme of a supervised adaptive filter. . . . .	43
4.2	Classification of adaptive filters based on the input-output relation. . . . .	46
5.1	A sparse acoustic impulse response. . . . .	61
6.1	Acoustic impulse responses used in simulations. . . . .	80

6.2	Misalignment of NLMS and IPNLMS algorithms with a white Gaussian noise input. . . . .	82
6.3	Misalignment of NLMS and IPNLMS algorithms with a female speech input. . . . .	83
6.4	ERLE of NLMS and IPNLMS algorithms with a female speech input. . . . .	84
6.5	Misalignment comparison of PAPA algorithms with different projection order. . . . .	84
6.6	Misalignment comparison between PAPA and PBAPA algorithms. . . . .	85
6.7	Misalignment comparison between NLMS and IPNLMS algorithms when a path change occurs. . . . .	86
6.8	ERLE comparison between NLMS and IPNLMS algorithms when the echo path changes. . . . .	87
6.9	Misalignment comparison between PAPA and PBAPA algorithms when the echo path changes. . . . .	88
6.10	Misalignment comparison between PAPA and VSS-PAPA algorithms with a white Gaussian noise input. . . . .	89
6.11	Misalignment comparison between PAPA and VSS-PAPA algorithms with a speech input. . . . .	90
6.12	ERLE comparison between PAPA and VSS-PAPA algorithms with a speech input. . . . .	91
6.13	Misalignment comparison between APA, PAPA and VSS-PAPA in presence of double talk. . . . .	92
6.14	Misalignment comparison between VSS-PAPA algorithms with and without a DTD in presence of double talk. . . . .	93
7.1	Echo transmission chain. . . . .	100
7.2	Frequency response of the loudspeaker APW300. . . . .	103
7.3	Comparison between the harmonic distortion of the loudspeaker APW300 and the neural loudspeaker model. . . . .	104
7.4	Distortion effect of the neural loudspeaker model. . . . .	105

---

7.5	Loss of quality caused by loudspeaker distortions in case of white Gaussian noise input. . . . .	106
7.6	Loss of quality caused by loudspeaker distortions in case of coloured noise input. . . . .	107
7.7	Loss of quality caused by loudspeaker distortions in case of speech input. . . . .	108
8.1	Hammerstein-based nonlinear system identification scheme. .	113
8.2	The functional link adaptive filter. . . . .	115
8.3	Functional link expansion. . . . .	117
8.4	Functional expansion in memoryless FLAF. . . . .	124
8.5	Functional expansion in FLAF with memory. . . . .	125
9.1	The split functional link adaptive filter. . . . .	133
9.2	Scheme of the nonlinearity introduced in the input signal. . .	135
9.3	(a) Far-end female speech input signal. (b) Signal acquired by the microphone after being distorted and reverberated. . . . .	136
9.4	Performance comparison in terms of ERLE between an FLAF and an SFLAF in case of white Gaussian input. . . . .	137
9.5	Performance comparison in terms of ERLE between an FLAF and an SFLAF in case of female speech input. . . . .	137
9.6	Performance comparison in terms of ERLE between the SFLAF and a 2nd order VF in case of white Gaussian input. . . . .	138
9.7	Performance comparison in terms of ERLE between the SFLAF and a 2nd order VF in case of female speech input. . . . .	139
9.8	Performance comparison in terms of ERLE between an SFLAF with memory and a memoryless SFLAF. . . . .	140
10.1	Possible variety of signals spanned by a certain degree of non-linearity and uncertainty. . . . .	147
10.2	Adaptive combination of transversal adaptive filters. . . . .	148
10.3	Convexity. . . . .	150

10.4	Convex combination of two adaptive filters. . . . .	151
11.1	Microphone array beamforming architecture. . . . .	160
11.2	Combined adaptive noise canceller architectures: (a) <i>system-by-system</i> and (b) <i>filter-by-filter</i> combination schemes. . . . .	164
11.3	Multi-stage combined adaptive noise canceller. . . . .	166
11.4	EMSE comparison between single-stage combined filtering architectures and conventional ones using the same projection order and different step size values. . . . .	169
11.5	Behaviour comparison between the mixing parameter of the system-by-system CANC and the mixing parameter relative to the first channel of the filter-by-filter CANC. . . . .	169
11.6	EMSE comparison between multi-stage combined filtering architectures and conventional ones. . . . .	171
11.7	EMSE comparison between multi-stage combined filtering architectures and single-stage ones. Multi-stage combined architecture always provide the best overall performance. . . . .	172
11.8	Speech enhancement nonstationary scenario. . . . .	173
12.1	Intelligent switching circuit . . . . .	179
12.2	Collaborative functional link adaptive filter. . . . .	181
12.3	Nonlinear adaptive path in a block-based collaborative FLAF. . . . .	184
12.4	Performance comparison in terms of ERLE between a linear, an SFLAF-based and a CFLAF-based echo cancellers in case of white Gaussian input. . . . .	187
12.5	Performance comparison in terms of ERLE between a linear, an SFLAF-based and a CFLAF-based echo cancellers in case of female speech input. . . . .	188

## LIST OF TABLES

7.1	Technical description of the loudspeaker model APW300, S.I.P.E. S.P.A. Electroacoustics. . . . .	102
11.1	SNR comparison in dB . . . . .	175





## LIST OF ACRONYMS

AEC	Acoustic Echo Cancellation
AI	Artificial Intelligence
AIR	Acoustic Impulse Response
ANC	Adaptive Noise Canceller
ANN	Artificial Neural Network
APA	Affine Projection Algorithm
AZK	All-Zero Kernel
BCFLAF	Block-Based Collaborative FLAF
BM	Blocking Matrix
CANC	Combined Adaptive Noise Canceller
CFLAF	Collaborative FLAF
DSB	Delay-and-Sum Beamformer

## LIST OF ACRONYMS

---

DSP	Digital Signal Processing
DTD	Double Talk Detector
EG $\pm$	Exponentiated Gradient
EMSE	Excess Mean Square Error
ERLE	Echo Return Loss Enhancement
FEB	Functional Expansion Block
FIR	Finite Impulse Response
FLAF	Functional Link Adaptive Filter
FLANN	Functional Link Artificial Neural Network
GSC	Generalized Sidelobe Canceller
IAI	Intelligent Acoustic Interface
IIR	Infinite Impulse Response
IPAPA	Improved PAPA
IPNLMS	Improved Proportionate NLMS
ISO	International Organization for Standardization
LMS	Least Mean Square
LSI	Linear Shift-Invariant
MIMO	Multiple-Input Multiple Output
MISO	Multiple-Input Single-Output
MLP	Multi-Layer Perceptron

MMSE	Minimum Mean Square Error
MSE	Mean Square Error
NAEC	Nonlinear Acoustic Echo Cancellation
NAPA	Natural Affine Projection Algorithm
NLMS	Normalized Least Mean Square
NNG	Normalized Natural Gradient
PAPA	Proportionate APA
PBAPA	Proportionate Block APA
PNLMS	Proportionate NLMS
QAM	Quadrature Amplitude Modulation
RBF	Radial Basis Function
RLS	Recursive Least Squares
SFLAF	Split FLAF
SISO	Single-Input Single-Output
SNR	Signal to Noise Ratio
ULA	Uniform Linear Array
VAD	Voice Activity Detection
VF	Volterra Filter
VSS	Variable Step Size