PART I

INTRODUCTION

—My work consists of two parts: of the one which is here,
and of everything which I have not written.
And precisely this second part is the important one.
Ludwig Wittgenstein

1

INTRODUCTION AND OUTLINE

Contents

1.1 MOTIVATIONS

Intelligent acoustic interfaces (IAIs) for hands-free speech communications are based on the modelling of acoustic paths and on the perception of complex sounds. In the development of such communication systems, many research areas intersect and cross-feed themselves, among which are: noise reduction, speech enhancement, acoustic echo cancellation, nonlinear channel modelling, multichannel acoustic modelling, source localization and tracking, blind source separation.

In such research context, matter of primary importance is the study of adaptive filtering algorithms and architectures [120]. Capabilities of such filtering structures to adapt to acoustic environments is that makes an acoustic interface intelligent. Moreover, adaptive filter performance bears on the quality of processed acoustic signals.

Among the foremost acoustic applications in which adaptive filtering plays a leading role are those on acoustic channel modelling, such as *acoustic echo cancellation* (AEC). The phenomenon of acoustic echo occurs when a delayed (and possibly distorted) version of the speech signal reproduced by a loudspeaker is acquired by a microphone and reflected back to remote user. An acoustic echo canceller aims at estimating the *acoustic impulse response* (AIR), i.e. modelling the acoustic path, in order to subtract the estimated echo signal from the microphone signal.

Therefore, the acoustic channel modelling represents an exhaustive issue in hands-free speech communications since it includes a set of problems common to the whole sector of acoustic scene analysis: the estimate of the impulse response, the presence of nonstationary elements in the environment, the presence of unwanted interfering signals, the presence of nonlinearities [12]. Such phenomena strongly degrade the perceived quality of the speech signal and might be tackled using signal processing techniques, that are pivotal in restoring the perceived intelligibility in a speech communication. This is the reason why the proposed research work mainly deals with applications on acoustic channel modelling, and in particular on AEC, in order to develop novel adaptive filtering techniques, which might also be used in other *distant-talking* applications.

Regarding the research in AEC, significant advances were achieved in the linear case, in which capabilities of adaptive filters have been exploited in order to model AIRs at best. In that sense significant results have been recently achieved for applications in hands-free speech communication in reverberant environments and in presence of interfering signals [12, 100], factors that cannot be neglected in immersive communications. However, similar results have not been reached yet in the nonlinear case.

The nonlinear case is characterized by the presence of distortions in the acoustic path that are funneled in the echo signal and cause a performance decrease and an even worse decrease of the perceived quality of information. Nonlinearities very often occur in acoustic applications since they are generated by loudspeakers or by the vibrations of audio devices' enclosures [147]. Therefore, nowadays, it is difficult to disregard echo cancellers that take into account nonlinearities, also due to a large spreading of low-cost audio devices, thus having low-quality electronic components and materials which may introduce even strong distortions.

Among the most popular nonlinear acoustic echo cancellers of recent years, stand out those based on adaptive Volterra filters [138, 23]. However, such nonlinear acoustic echo cancellers involve computational costs that are definitely larger than conventional echo cancellers (i.e. linear echo cancellers) and, moreover, they may provide worse performance compared to the last ones. That affects also the strategies of many companies that provide teleconferencing services, which often choose to drop the use of nonlinear echo cancellers even at the expense of communication quality. On the other side, these are also the main motivations that underpin the proposed research project.

1.2 SCOPE OF THE WORK

The development of adaptive algorithms for intelligent acoustic interfaces is based on high-complexity scenarios which take into account several phenomena that may degrade the speech intelligibility in a hands-free speech communication. We start from an analysis of such interfering phenomena that may be essentially labelled as linear or nonlinear events. Such division allows to design *ad hoc* adaptive algorithms, thus making acoustic intelligent interfaces robust against interfering signals.

Regarding the acoustic channel modelling in the linear case it is sufficient to investigate about adaptive models that are statistically robust. However, in order to recreate accurately an acoustic scene free from any interference, noise and unwanted signals, it is advisable to perform a nonlinear processing of acquired information that is able to learn from the environment in a supervised or unsupervised way. In both the cases, linear and nonlinear, automatic learning and continuous adaptivity are fundamental elements to satisfy quality requirements of speech communication [12].

In order to tackle linear interfering signals, we deal with a recently proposed filtering technique that is based on *proportionate adaptive filters* [100]. This family of algorithms exploits sparsity constraints that are typical of AIRs, thus yielding a performance improvement which is able to reduce the limits posed by acoustic environments. The investigation about such family of algorithms bears to the formulation of a framework for the derivation of (linear) adaptive filters and to the development of efficient proportionate adaptive algorithms for immersive speech communication.

On the other side, in order to tackle nonlinearities in acoustic channel modelling, we propose a novel nonlinear filtering model based on *functional links*. From such nonlinear model we develop some algorithms and architectures on purpose of *nonlinear acoustic echo cancellation* (NAEC). The main idea which underpins such *functional link adaptive filters* is that of estimating and modelling nonlinearities introduced in the echo path by the environment and interfering sources, and then cancelling them, thus improving the perceived quality of acoustic information.

Moreover, both in linear and nonlinear cases, the proposed adaptive algorithms are used to form more complex filtering architectures based on the *adaptive combination of filters*. Such architectures result more robust against several kinds of adverse environment conditions compared to conventional filtering techniques.

1.3 ORGANIZATION

The proposed research project is structured in three main parts: the first one dealing with linear adaptive algorithms, the second one with nonlinear adaptive algorithms and the last one dealing with robust filtering architectures. An introducing part is also added at the beginning of the work, as well as a conclusive part is added at the end. In detail, this dissertation is organized as follows:

Part I introduces some preliminary basics.

- **Chapter 1** describes the motivation and the scope of our proposal.
- **Chapter 2** introduces intelligent acoustic interfaces and their application in immersive speech communications.
- **Chapter 3** explains the formulation of main problems in hands-free speech communications that we aim at tackling with adaptive algorithms.

Part II deals with adaptive algorithms designed to address those problems classified as linear.

- **Chapter 4** introduces a brief view on the theory of adaptive filtering.
- **Chapter 5** introduces proportionate adaptive algorithms according to the proposed general framework.
- **Chapter 6** describes by means of simulations the most important features of the proportionate adaptive algorithms introduced in the previous chapter.

Part III deals with adaptive algorithms designed to tackle the presence of nonlinearities in the acoustic channel.

- **Chapter 7** formulates the problem of nonlinearities which cause an important limitation to the achievable speech quality.

Chapter 8 introduces a new class of nonlinear algorithms, the functional link adaptive filters, whose structure is based on Hammerstein model.

Chapter 9 describes some variants of functional link adaptive filters properly designed for nonlinear acoustic echo cancellation.

Part IV introduces more complex architectures based on adaptive combination of filters to increase robustness against adverse acoustic environments.

Chapter 10 introduces intelligent circuits based on the adaptive combination of filters.

Chapter 11 describes combined architectures for speech enhancement in multisource environments.

Chapter 12 describes collaborative architectures for nonlinear acoustic echo cancellation.

Part V draws our conclusions.

Chapter 13 concludes the work and introduces possible future perspectives.

1.4 NOTATION

In this dissertation, matrices are represented by boldface capital letters and vectors are denoted by boldface lowercase letters. Time-varying vectors and matrices show discrete-time index as a subscript index, while in time-varying scalar elements the time index is denoted in square brackets. A regression vector is represented as $\mathbf{x}_n \in \mathbb{R}^M = \left[\begin{array}{cccc} x[n] & x[n-1] & \dots & x[n-M+1] \end{array} \right]^T$, where M is the overall vector length and $x[n-i]$ is individual entry at the generic time instant $n-i$. On the other side, a snap-shot vector, which includes a number Q of different contributions at n-th time instant, is represented as

$\mathbf{y}[n] \in \mathbb{R}^Q = \begin{bmatrix} y_0[n] & y_1[n] & \dots & y_{Q-1}[n] \end{bmatrix}^T$. However, a generic coefficient vector, in which all elements depend on the same time instant, is denoted as $\mathbf{w}_n \in \mathbb{R}^M = \begin{bmatrix} w_0[n] & w_1[n] & \dots & w_{M-1}[n] \end{bmatrix}^T$, where $w_i[n]$ is the generic i-th individual entry at n-th time instant. When the coefficient vector is a realization of a time-invariant process the time index is omitted. All vectors are represented as column vectors.