# A New Decision-Theory-Based Framework for Echo Canceler Control

Tales Imbiriba,Jose´ Carlos M. Bermudez, Jean-Yves Tourneret and Neil J. Bershad

***Abstract*—A control logic has a central role in many echo cancellation systems for optimizing the performance of adaptive filters, while estimating the echo path. For reliable control, accurate double-talk and channel change detectors are usually incorporated to the echo canceler. This work expands the usual detection strategy to define a classification problem characterizing four possible states of the echo canceler operation. The new formulation allows the use of decision theory to continuously control the transitions among the different modes of operation. The classification rule reduces to a low-cost statistics, for which it is possible to determine the probability of error under all hypotheses, allowing the classification performance to be accessed analytically. Monte Carlo simulations using synthetic and real data illustrate the reliability of the proposed method.**

***Index Terms*—Adaptive filters, adaptive signal processing, adaptive systems, echo cancellation, channel change, double-talk, classification, multivariate gamma distribution.**

## I. Introduction

ECHO cancellation is a requirement in modern voice communication systems. Speech echo cancelers (ECs) are employed in telephone networks (line echo cancelers) or in hands-free communications (acoustic echo cancelers). Most EC designs include two main blocks; a channel identification block and a control logic block. The channel identification block tries to estimate the echo path, often employing adaptive filtering. However, the adaptive algorithm tends to diverge in the presence of near-end signals (double-talk – DT). Hence, adaptation must be stopped during DT. On the other hand, abrupt echo channel changes (CC) require a faster adaptation to improve tracking. Finally, in the absence of both DT and CC, a slow adaptation rate tends to improve channel estimation accuracy. The control logic is then required to control the transitions among these distinct modes of adaptive operation.

The EC control may or may not employ DT or CC detectors. Different approaches have been proposed to deal with DT or CC in echo cancelers, some of which do not require a DT detector, aiming at a continuous adaptation of the EC. Blind source separation strategies based on independent component analysis (BSS/ICA) were proposed in [1], [2], variable stepsize (VSS) methods in [3]–[5], and methods based on the prediction of the near-end signal using a prediction error (PE) framework in [4], [6], [7]. Frequency domain adaptive filter (FDAF) solutions have also been proposed, resulting in low computational complexity and fast convergence at the expense of higher memory usage and additional end-to-end delay [4], [5], [8].

BSS strategies try to separate the near- and far-end signal components, adapting the EC only on the far-end component. A BSS method, based on ICA, led to a weighted recursive least-squares (RLS) [1], [2] algorithm using a $O(N^2)$ implementation based on the matrix inversion lemma to guarantee stability. VSS methods continuously adjust the adaptive filter (AF) stepsize to cope with different states of a dynamical system. An optimal AF stepsize is derived in [3] which depends on the non-accessible undisturbed error signal. This requires further estimation and detection stages. VSS strategies were also considered in the frequency domain [4], [5]. An FDAF VSS method was proposed in [4] in the context of PE framework. The gains of a noise-reduction Wiener filter where used as variable stepsizes for each frequency bin in the FDAF. In [5], the FDAF at each frequency bin is derived from a system distance measure as a function of time and frequency.

The PE framework has been used in [4], [6], [7] to simultaneously estimate the echo path and a parametric AR model for the near-end signal. A low complexity method was proposed for the near-end modeling that is adequate for speech signals, leading to the PEM-AFROW algorithm [6], [7]. Although the resulting cost function is essentially nonconvex, simulations indicate that the proposed algorithms are robust to double-talk and present fast convergence under single-talk.

Several works have proposed methods for DT detection in ECs without considerations regarding CC, such as [9]–[12]. However, DT detection strategies that assume a static channel

This work was supported by the Brazilian National Council for Sci-entific and Technological Development under Grant 304250/2017-1 and Grant 43857/2016-9.

T. Imbiriba and J. C. M. Bermudez are with the Department of Electrical Engineering, Federal University of Santa Catarina, Florianópolis 88040-900, Brazil (e-mail: talesim@gmail.com; j.bermudez@ieee.org).

J.-Y. Tourneret is with the Institut de Recherche en Informatique de Toulouse, Ecole Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique et d'Hydraulique de Toulouse, Télécommunications and Space for Aeronautics (TéSA), University of Toulouse, 31071 Toulouse, France (e-mail: Jean-Yves.Tourneret@enseeiht.fr).

N. J. Bershad is with the Department of Electrical Engineering and Computer Science, University of California Irvine, Irvine, CA 92660 USA (e-mail: bershad@ece.uci.edu).

Digital Object Identifier 10.1109/TSP.2018.2849748

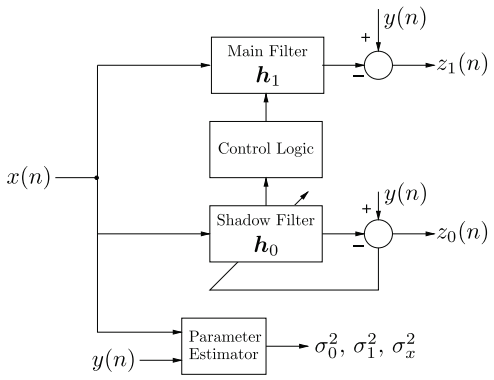


Fig. 1. Basic echo canceller structure.

response may yield unpredictable performances in the presence of CC [13]. The vast majority of the techniques available for DT or CC detection rely on ad hoc statistics to make the decision, leading to cumbersome design processes. A few works employ a statistical framework to formulate the detection problem. For instance, [14] proposes a maximum a posteriori (MAP) decision rule based on channel output observations and assuming Bernoulli distributed priors for the different hypotheses. A similar approach is used in [15], but employing a Markov channel model. In [16], a generalized likelihood ratio test (GLRT) is proposed using observations from both the channel input and output signals. DT and CC detection are considered. In [17] and [18], a first test distinguishes single-talk from DT or CC, and a second test based on the echo path estimate detects DT. Though the latter two studies consider DT and CC in a single formulation, all these aforementioned statistical formulations have been proposed for the conventional adaptive EC structure [3].

An alternative EC structure has been proposed in [9], which uses a shadow adaptive filter that operates in parallel with the actual echo cancellation filter. The shadow filter coefficients are transferred to the echo cancellation filter when the shadow filter is a better estimate of the unknown channel response than the echo cancellation filter. From the authors experience, this structure allows a much better control of the EC convergence than the conventional structure. The EC structure is shown in Fig. 1. The EC consists of the main echo cancellation filter and the adaptive shadow filter. The output of the main filter is subtracted from the echo to obtain the canceled echo $z_1(n)$. The shadow filter weights are adapted continuously. The control logic is designed such that the shadow filter coefficients are copied to the main filter when this will improve the EC performance. A likelihood ratio test (LRT) detector based on the EC structure in Fig. 1 was derived in [19] to detect DT versus CC. A generalized LRT (GLRT) that could be simplified to a sufficient statistic was proposed for the same EC structure in [20]. The performance of the test statistic was evaluated as a function of the system parameters. The idea developed in [19] and [20] was to use the detection result 1) to stop adaptation when DT was detected and 2) to adapt fast in the presence of channel change. The speed of adaptation was controlled by the adaptation stepsize.

The decision theory-based DT and CC detection formulation in [19], [20] did not include decision theory based formulations for the exit from a DT or a CC condition. These decisions were still made in an ad hoc manner.

This paper formulates the echo canceler control logic as a more general classification problem, with four hypotheses associated to the presence or absence of DT and to the presence or absence of CC[1]

$$\begin{aligned}\mathcal{H}_0 &: \text{no DT and no CC}\\ \mathcal{H}_1 &: \text{no DT and CC}\\ \mathcal{H}_2 &: \text{DT and no CC}\\ \mathcal{H}_3 &: \text{DT and CC.}\end{aligned} \tag{1}$$

There are several motivations for identifying these four classes. These motivations include 1) the possibility to adjust accurately the stepsize of the adaptive filter for long time intervals when there is no DT and no CC, resulting in smaller residual errors, 2) the inclusion of $\mathcal{H}_3$ adds an important degree of flexibility to the control logic that can be exploited, as will be shown in Section V-A, 3) these four classes lead to a simple and low cost test statistic.

The paper is organized as follows. In Section II we introduce the signal models and derive the classification rules. In Section III we present the performance analysis of the proposed classifier. Monte Carlo simulations are presented in Section IV to validate the theory. Section V discusses application of the proposed classification strategy and presents illustrative simulation results. Finally, Section VI discusses the results and presents the conclusions.

## II. Double-Talk and Channel Change Classification

### A. Signal and Channel Models

The channel input vector $\boldsymbol{x}(n) = [x(n), \ldots, x(n-N+1)]^\top$ is of dimension $N \times 1$ with covariance matrix $E[\boldsymbol{x}(n)\boldsymbol{x}^\top(n)] = \boldsymbol{\Sigma}_x$ and the channel output is a scalar $y(n)$. The input signal is stationary within the decision periods and the DT signal can be modelled by a white Gaussian process for detection purposes [16]. Also, $[y(n), \boldsymbol{x}^\top(n)]^\top$ is modelled as a zero-mean Gaussian vector. Denoting the adaptive shadow filter response by $\boldsymbol{h}_0$, the main echo cancellation filter response by $\boldsymbol{h}_1$, and the true echo path response by $\boldsymbol{g}$, the channel output $y(n)$ can be expressed as follows under the different hypotheses:

$$\begin{aligned}&\mathcal{H}_0(\text{no DT, no CC}) : \boldsymbol{h}_1 = \boldsymbol{g},\ y(n) = \boldsymbol{h}_1^\top \boldsymbol{x}(n) + n_0(n)\\ &\mathcal{H}_1(\text{no DT, CC}) : \boldsymbol{h}_0 = \boldsymbol{g},\ y(n) = \boldsymbol{h}_0^\top \boldsymbol{x}(n) + n_0(n)\\ &\mathcal{H}_2(\text{DT, no CC}) : \boldsymbol{h}_1 = \boldsymbol{g},\ y(n) = \boldsymbol{h}_1^\top \boldsymbol{x}(n) + n_0(n) + n_1(n)\\ &\mathcal{H}_3(\text{DT, CC}) : \boldsymbol{h}_0 = \boldsymbol{g},\ y(n) = \boldsymbol{h}_0^\top \boldsymbol{x}(n) + n_0(n) + n_1(n).\end{aligned} \tag{2}$$

[1]The acronyms CC and DT are usually employed to signify instantaneous phenomena. Here, CC and DT are used to define system states following the onset of channel change or double-talk, which are tested for at regular time intervals (see Section V).

The $\mathcal{H}_0$ hypothesis considers that $\boldsymbol{h}_0$ has converged and has been recently copied to $\boldsymbol{h}_1$. Hypothesis $\mathcal{H}_1$ assumes that $\boldsymbol{h}_0$ has already converged (or is much closer to $\boldsymbol{g}$ than $\boldsymbol{h}_1$) after a channel change. Therefore, we consider that the system is at a CC state when $\boldsymbol{h}_0 \approx \boldsymbol{g}$ and there exist a measurable mismatch between $\boldsymbol{h}_0$ and $\boldsymbol{h}_1$. In $\mathcal{H}_2$, a DT signal $n_1(n)$ happens after convergence of $\boldsymbol{h}_0$ and copy to $\boldsymbol{h}_1$ (similar to $\mathcal{H}_0$). Finally, a fourth hypothesis $\mathcal{H}_3$ considers that DT happens following a CC after $\boldsymbol{h}_0$ has already converged to the new channel but has not yet been copied to $\boldsymbol{h}_1$. All cases rely on the convergence (or divergence) of $\boldsymbol{h}_0$ and its relation to $\boldsymbol{h}_1$ resulting in several practical implications concerning the control logic block in Fig. 1. The control strategy will be addressed in Section V.

The additive noise $n_0(n)$ is stationary zero-mean white[2] Gaussian, independent of $\boldsymbol{x}(n)$ with $E[n_0^2(n)] = \sigma_0^2$. The second additive noise $n_1(n)$, modeling the DT, is zero-mean white Gaussian, and independent of both $\boldsymbol{x}(n)$ and $n_0(n)$ with $E[n_1^2(n)] = \sigma_1^2$. Two error signals $z_0(n) = y(n) - \boldsymbol{h}_0^\top \boldsymbol{x}(n)$ and $z_1(n) = y(n) - \boldsymbol{h}_1^\top \boldsymbol{x}(n)$ were introduced in [20] to facilitate the analysis. These error signals can be expressed as follows under the different hypotheses

$$
\begin{aligned}
&\mathcal{H}_0 \text{(no DT, no CC)}: \\
&\quad z_0(n) = (\boldsymbol{h}_1 - \boldsymbol{h}_0)^\top \boldsymbol{x}(n) + n_0(n),\ z_1(n) = n_0(n) \\
&\mathcal{H}_1 \text{(no DT, CC)}: \\
&\quad z_0(n) = n_0(n),\ z_1(n) = (\boldsymbol{h}_0 - \boldsymbol{h}_1)^\top \boldsymbol{x}(n) + n_0(n) \\
&\mathcal{H}_2 \text{(DT, no CC)}: \\
&\quad z_0(n) = (\boldsymbol{h}_1 - \boldsymbol{h}_0)^\top \boldsymbol{x}(n) + n_0(n) + n_1(n) \\
&\quad z_1(n) = n_0(n) + n_1(n) \\
&\mathcal{H}_3 \text{(DT, CC)}: \\
&\quad z_0(n) = n_0(n) + n_1(n) \\
&\quad z_1(n) = (\boldsymbol{h}_0 - \boldsymbol{h}_1)^\top \boldsymbol{x}(n) + n_0(n) + n_1(n).
\end{aligned} \tag{3}
$$

### B. Classification Rule

*1) One-Sample Case:* The joint pdf of $\boldsymbol{z}(n) = [z_0(n), z_1(n)]^\top$ is Gaussian under all hypotheses such that

$$
p[\boldsymbol{z}(n)|\mathcal{H}_i] \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{i1}), \quad i = 0, \ldots, 3 \tag{4}
$$

where the second subscript in $\boldsymbol{\Sigma}_{i1}$ (1 in this case) indicates the 1-sample case. The covariance matrices of $\boldsymbol{z}(n)$ under the different hypotheses can be written

$$
\begin{aligned}
\boldsymbol{\Sigma}_{01} &= \begin{pmatrix} \sigma_0^2 + c_x^2 & \sigma_0^2 \\ \sigma_0^2 & \sigma_0^2 \end{pmatrix} \\
\boldsymbol{\Sigma}_{11} &= \begin{pmatrix} \sigma_0^2 & \sigma_0^2 \\ \sigma_0^2 & \sigma_0^2 + c_x^2 \end{pmatrix}
\end{aligned} \tag{5}
$$

[2]Note here that the whiteness assumption for $n_0(n)$ is not restrictive since it is always possible to whiten the channel outputs by pre-multiplying consecutive samples by an appropriate matrix. Of course, this operation assumes that the covariance matrix of consecutive noise samples is known or can be estimated.

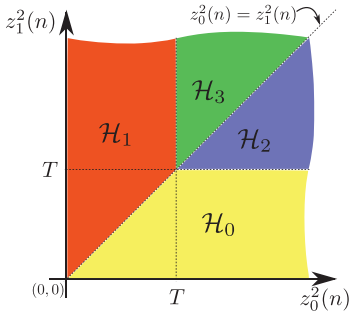


Fig. 2. DT and CC decision regions in the $(z_0^2(n), z_1^2(n))$ plane.

$$
\begin{aligned}
\boldsymbol{\Sigma}_{21} &= \begin{pmatrix} \sigma_0^2 + \sigma_1^2 + c_x^2 & \sigma_0^2 + \sigma_1^2 \\ \sigma_0^2 + \sigma_1^2 & \sigma_0^2 + \sigma_1^2 \end{pmatrix} \\
\boldsymbol{\Sigma}_{31} &= \begin{pmatrix} \sigma_0^2 + \sigma_1^2 & \sigma_0^2 + \sigma_1^2 \\ \sigma_0^2 + \sigma_1^2 & \sigma_0^2 + \sigma_1^2 + c_x^2 \end{pmatrix}
\end{aligned} \tag{6}
$$

with

$$
c_x^2 = (\boldsymbol{h}_0 - \boldsymbol{h}_1)^\top \boldsymbol{\Sigma}_x (\boldsymbol{h}_0 - \boldsymbol{h}_1) \tag{7}
$$

where $c_x^2$ can be interpreted as the power at the output of the difference filter with response $\boldsymbol{h}_0 - \boldsymbol{h}_1$. Assuming all hypotheses are equiprobable, the classification rule minimizing the average probability of error decides hypothesis $\mathcal{H}_i$ is true when

$$
\begin{aligned}
&\frac{1}{\sqrt{|\boldsymbol{\Sigma}_{i1}|}} \exp\left[-\frac{1}{2}\boldsymbol{z}^\top(n)\boldsymbol{\Sigma}_{i1}^{-1}\boldsymbol{z}(n)\right] \\
&\qquad > \frac{1}{\sqrt{|\boldsymbol{\Sigma}_{j1}|}} \exp\left[-\frac{1}{2}\boldsymbol{z}^\top(n)\boldsymbol{\Sigma}_{j1}^{-1}\boldsymbol{z}(n)\right]
\end{aligned} \tag{8}
$$

for all $j \neq i$. Equivalently, hypothesis $\mathcal{H}_i$ will be accepted if

$$
\boldsymbol{z}^\top(n)\left(\boldsymbol{\Sigma}_{j1}^{-1} - \boldsymbol{\Sigma}_{i1}^{-1}\right)\boldsymbol{z}(n) > \ln\left(\frac{|\boldsymbol{\Sigma}_{i1}|}{|\boldsymbol{\Sigma}_{j1}|}\right) \tag{9}
$$

for all $j \neq i$. Straightforward computations (detailed in Appendix A) allow one to compute the inverses and determinants of the $2 \times 2$ matrices $\boldsymbol{\Sigma}_{i1}$ and $\boldsymbol{\Sigma}_{j1}$ yielding the following classification rule

$$
\begin{aligned}
&\mathcal{H}_0 \text{ aif } z_1^2(n) < z_0^2(n) \text{ and } z_1^2(n) < T \\
&\mathcal{H}_1 \text{ aif } z_1^2(n) > z_0^2(n) \text{ and } z_0^2(n) < T \\
&\mathcal{H}_2 \text{ aif } z_1^2(n) < z_0^2(n) \text{ and } z_1^2(n) > T \\
&\mathcal{H}_3 \text{ aif } z_1^2(n) > z_0^2(n) \text{ and } z_0^2(n) > T
\end{aligned} \tag{10}
$$

where "aif" means "accepted if" and

$$
T = \frac{\sigma_0^2(\sigma_0^2 + \sigma_1^2)}{\sigma_1^2} \ln\left(1 + \frac{\sigma_1^2}{\sigma_0^2}\right). \tag{11}
$$

The different decision regions corresponding to (10) are illustrated in the $(z_0^2(n), z_1^2(n))$ plane in Fig. 2.

*2) Multiple Samples:* The analysis above can be generalized to the case where multiple time samples $\boldsymbol{z}(n-k)$, for $k = p-1, \ldots, 0$, are available. The analysis is performed here for two samples (i.e., $p = 2$) for simplicity and is generalized later.

When two samples are observed, the error signals $z_0(n)$, $z_0(n-1)$ and $z_1(n)$, $z_1(n-1)$ are considered. They can be expressed as follows under the different hypotheses:

Under $H_0$:

$$\begin{aligned} z_0(n) &= (\boldsymbol{h}_1 - \boldsymbol{h}_0)^\top \boldsymbol{x}(n) + n_0(n) \\ z_1(n) &= n_0(n) \\ z_0(n-1) &= (\boldsymbol{h}_1 - \boldsymbol{h}_0)^\top \boldsymbol{x}(n-1) + n_0(n-1) \\ z_1(n-1) &= n_0(n-1). \end{aligned} \tag{12}$$

Under $H_1$:

$$\begin{aligned} z_0(n) &= n_0(n) \\ z_1(n) &= (\boldsymbol{h}_0 - \boldsymbol{h}_1)^\top \boldsymbol{x}(n) + n_0(n) \\ z_0(n-1) &= n_0(n-1) \\ z_1(n-1) &= (\boldsymbol{h}_0 - \boldsymbol{h}_1)^\top \boldsymbol{x}(n-1) + n_0(n-1). \end{aligned} \tag{13}$$

Under $H_2$:

$$\begin{aligned} z_0(n) &= (\boldsymbol{h}_1 - \boldsymbol{h}_0)^\top \boldsymbol{x}(n) + n_0(n) + n_1(n) \\ z_1(n) &= n_0(n) + n_1(n) \\ z_0(n-1) &= (\boldsymbol{h}_1 - \boldsymbol{h}_0)^\top \boldsymbol{x}(n-1) \\ &\quad + n_0(n-1) + n_1(n-1) \\ z_1(n-1) &= n_0(n-1) + n_1(n-1). \end{aligned} \tag{14}$$

Under $H_3$:

$$\begin{aligned} z_0(n) &= n_0(n) + n_1(n) \\ z_1(n) &= (\boldsymbol{h}_0 - \boldsymbol{h}_1)^\top \boldsymbol{x}(n) + n_0(n) + n_1(n) \\ z_0(n-1) &= n_0(n-1) + n_1(n-1) \\ z_1(n-1) &= (\boldsymbol{h}_0 - \boldsymbol{h}_1)^\top \boldsymbol{x}(n-1) \\ &\quad + n_0(n-1) + n_1(n-1). \end{aligned} \tag{15}$$

Defining $\boldsymbol{z}_{2d}(n) = [z_0(n), z_0(n-1), z_1(n), z_1(n-1)]^\top$, $\boldsymbol{z}_{2d}(n)$ is a zero-mean Gaussian vector under all hypotheses. Straightforward computations yield the covariance matrices of $\boldsymbol{z}_{2d}(n)$ under the different hypotheses. These matrices can be expressed as

$$\begin{aligned} \boldsymbol{\Sigma}_{02} &= \begin{pmatrix} \sigma_0^2 \boldsymbol{I}_2 + \boldsymbol{H}_x & \sigma_0^2 \boldsymbol{I}_2 \\ \sigma_0^2 \boldsymbol{I}_2 & \sigma_0^2 \boldsymbol{I}_2 \end{pmatrix} \\ \boldsymbol{\Sigma}_{12} &= \begin{pmatrix} \sigma_0^2 \boldsymbol{I}_2 & \sigma_0^2 \boldsymbol{I}_2 \\ \sigma_0^2 \boldsymbol{I}_2 & \sigma_0^2 \boldsymbol{I}_2 + \boldsymbol{H}_x \end{pmatrix} \\ \boldsymbol{\Sigma}_{22} &= \begin{pmatrix} (\sigma_0^2 + \sigma_1^2) \boldsymbol{I}_2 + \boldsymbol{H}_x & (\sigma_0^2 + \sigma_1^2) \boldsymbol{I}_2 \\ (\sigma_0^2 + \sigma_1^2) \boldsymbol{I}_2 & (\sigma_0^2 + \sigma_1^2) \boldsymbol{I}_2 \end{pmatrix} \end{aligned} \tag{16}$$

$$\boldsymbol{\Sigma}_{32} = \begin{pmatrix} (\sigma_0^2 + \sigma_1^2) \boldsymbol{I}_2 & (\sigma_0^2 + \sigma_1^2) \boldsymbol{I}_2 \\ (\sigma_0^2 + \sigma_1^2) \boldsymbol{I}_2 & (\sigma_0^2 + \sigma_1^2) \boldsymbol{I}_2 + \boldsymbol{H}_x \end{pmatrix} \tag{17}$$

where $\boldsymbol{I}_2$ is the $2 \times 2$ identity matrix and $\boldsymbol{H}_x$ is given by Equation (18).

$$\begin{aligned} \boldsymbol{H}_x = &\begin{pmatrix} \boldsymbol{h}_0 - \boldsymbol{h}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{h}_0 - \boldsymbol{h}_1 \end{pmatrix}^\top \begin{pmatrix} \boldsymbol{\Sigma}_x & \boldsymbol{R}_{1x} \\ \boldsymbol{R}_{-1x} & \boldsymbol{\Sigma}_x \end{pmatrix} \\ &\times \begin{pmatrix} \boldsymbol{h}_0 - \boldsymbol{h}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{h}_0 - \boldsymbol{h}_1 \end{pmatrix} \end{aligned} \tag{18}$$

In (18), $\boldsymbol{\Sigma}_x = E[\boldsymbol{x}(n)\boldsymbol{x}^\top(n)]$, $\boldsymbol{R}_{1x} = E[\boldsymbol{x}(n)\boldsymbol{x}^\top(n-1)]$, and $\boldsymbol{R}_{-1x} = E[\boldsymbol{x}(n-1)\boldsymbol{x}^\top(n)]$. The determinants and inverses of these block matrices can be computed following [21, p. 572]

$$\begin{aligned} |\boldsymbol{\Sigma}_{02}| &= |\boldsymbol{\Sigma}_{12}| = \sigma_0^4 |\boldsymbol{H}_x| \\ |\boldsymbol{\Sigma}_{22}| &= |\boldsymbol{\Sigma}_{32}| = (\sigma_0^2 + \sigma_1^2)^2 |\boldsymbol{H}_x| \end{aligned} \tag{19}$$

and

$$\begin{aligned} \boldsymbol{\Sigma}_{02}^{-1} &= \begin{pmatrix} \boldsymbol{H}_x^{-1} & -\boldsymbol{H}_x^{-1} \\ -\boldsymbol{H}_x^{-1} & \frac{1}{\sigma_0^2} \boldsymbol{I}_2 + \boldsymbol{H}_x^{-1} \end{pmatrix} \\ \boldsymbol{\Sigma}_{12}^{-1} &= \begin{pmatrix} \frac{1}{\sigma_0^2} \boldsymbol{I}_2 + \boldsymbol{H}_x^{-1} & -\boldsymbol{H}_x^{-1} \\ -\boldsymbol{H}_x^{-1} & \boldsymbol{H}_x^{-1} \end{pmatrix} \\ \boldsymbol{\Sigma}_{22}^{-1} &= \begin{pmatrix} \boldsymbol{H}_x^{-1} & -\boldsymbol{H}_x^{-1} \\ -\boldsymbol{H}_x^{-1} & \frac{1}{\sigma_0^2 + \sigma_1^2} \boldsymbol{I}_2 + \boldsymbol{H}_x^{-1} \end{pmatrix} \end{aligned} \tag{20}$$

$$\boldsymbol{\Sigma}_{32}^{-1} = \begin{pmatrix} \frac{1}{\sigma_0^2 + \sigma_1^2} \boldsymbol{I}_2 + \boldsymbol{H}_x^{-1} & -\boldsymbol{H}_x^{-1} \\ -\boldsymbol{H}_x^{-1} & \boldsymbol{H}_x^{-1} \end{pmatrix}. \tag{21}$$

where $\boldsymbol{H}_x^{-1}$ is assumed to exist.

Performing the same computations shown in Appendix A for vector $\boldsymbol{z}_{2d}(n)$ and matrices (16) and (17), the following multiple sample classification rule can then be obtained

$$\begin{aligned} &\mathcal{H}_0 \text{ aif } \|\boldsymbol{z}_1(n)\|^2 < \|\boldsymbol{z}_0(n)\|^2 \text{ and } \|\boldsymbol{z}_1(n)\|^2 < T_2, \\ &\mathcal{H}_1 \text{ aif } \|\boldsymbol{z}_1(n)\|^2 > \|\boldsymbol{z}_0(n)\|^2 \text{ and } \|\boldsymbol{z}_0(n)\|^2 < T_2, \\ &\mathcal{H}_2 \text{ aif } \|\boldsymbol{z}_1(n)\|^2 < \|\boldsymbol{z}_0(n)\|^2 \text{ and } \|\boldsymbol{z}_1(n)\|^2 > T_2, \\ &\mathcal{H}_3 \text{ aif } \|\boldsymbol{z}_1(n)\|^2 > \|\boldsymbol{z}_0(n)\|^2 \text{ and } \|\boldsymbol{z}_0(n)\|^2 > T_2, \end{aligned} \tag{22}$$

where $\boldsymbol{z}_i(\,n\,) = [z_i(\,n\,), z_i(\,n-1\,)]^\top$, $\|\boldsymbol{z}_i(n)\|^2 = z_i^2(n) + z_i^2(n-1)$ and

$$T_2 = 2T = 2 \frac{\sigma_0^2 \left(\sigma_0^2 + \sigma_1^2\right)}{\sigma_1^2} \ln\left(1 + \frac{\sigma_1^2}{\sigma_0^2}\right). \tag{23}$$

The factor 2 multiplying $T$ in (23) results from $\ln\left(|\boldsymbol{\Sigma}_{i1}|/|\boldsymbol{\Sigma}_{j1}|\right) = -2\ln\left(1 + \sigma_1^2/\sigma_0^2\right)$. This result can be compared with (10) obtained for the one-sample case. The generalization to more than two samples is straightforward. Indeed, in the $p$-sample case, the covariance matrices $\boldsymbol{\Sigma}_{ip}$ of $\boldsymbol{z}_{pd}(n)$ are defined as in (16) and (17), with $\boldsymbol{I}_2$ replaced with $\boldsymbol{I}_p$, and $\boldsymbol{H}_x$ defined differently. However, since $\boldsymbol{H}_x$ cancels from the difference between the two inverses, the classification rule for the $p$-sample case is expressed by (22) with $\|\boldsymbol{z}_i(n)\|^2 = \boldsymbol{z}_i^\top(n)\boldsymbol{z}_i(n) = \sum_{k=0}^{p-1} z_i^2(n-k)$ the squared norm of $\boldsymbol{z}_i(n)$, $i = 0, 1$, and with $T_2 = 2T$ replaced with $T_p = pT$.

## III. Performance Analysis

This section studies the probability of classification error for the classifier proposed in Section II.

### A. One-Sample Case

It is clear from the classification rules (10) that $\boldsymbol{d}(n) = [z_0^2(n), z_1^2(n)]^\top$ is a sufficient statistic for the classification problem. Interestingly, the exact distribution of $\boldsymbol{d}(n)$ can be derived under all hypotheses, allowing for an analytical study of the classifier performance. First, we note that the elements of $\boldsymbol{d}(n)$ form the diagonal of the matrix $\boldsymbol{Z} = \boldsymbol{z}(n)\boldsymbol{z}^\top(n)$. Now, since $\boldsymbol{z}(n) = [z_0(n), z_1(n)]^\top$ is jointly distributed according to a zero-mean Gaussian distribution with covariance matrix $\boldsymbol{\Sigma}_{i1}$, see (4), it is shown in Appendix B that, under all hypotheses $\mathcal{H}_i$, $i = 0, \ldots, 3$, $\boldsymbol{d}(n)$ is distributed according to a multivariate gamma distribution denoted $\mathcal{G}(q, P)$ with shape parameter $q = p/2$ and scale parameter $P = \{p_1, p_2, p_{12}\}$, with

$$\begin{aligned} p_1 &= 2\boldsymbol{\Sigma}_{i1}(1,1) \\ p_2 &= 2\boldsymbol{\Sigma}_{i1}(2,2) \\ p_{12} &= 4\left[\boldsymbol{\Sigma}_{i1}(1,1)\boldsymbol{\Sigma}_{i,1}(2,2) - \boldsymbol{\Sigma}_{i1}(1,2)\boldsymbol{\Sigma}_{i1}(2,1)\right] \end{aligned} \quad (24)$$

where $\boldsymbol{\Sigma}_{i1}(1,1)$, $\boldsymbol{\Sigma}_{i1}(1,2) = \boldsymbol{\Sigma}_{i1}(2,1)$ and $\boldsymbol{\Sigma}_{i1}(2,2)$ are the elements of the covariance matrix $\boldsymbol{\Sigma}_{i1}$.

### B. Multiple-Sample

Once again it is clear that the vector $\boldsymbol{d}(n) = [\|\boldsymbol{z}_0(n)\|^2, \|\boldsymbol{z}_1(n)\|^2]^\top$ is a sufficient statistic for solving the proposed classification problem. Noting that $\boldsymbol{z}_{pd}(n)$ is a rearrangement of the $p$ vectors $\boldsymbol{z}(n-k)$, $k = 0, \ldots, p-1$, the distribution of $\boldsymbol{d}(n)$ can be obtained following the reasoning presented in Appendix B, under the assumption of independence of vectors $\boldsymbol{z}(n-i)$ and $\boldsymbol{z}(n-j)$, $i \neq j$, and stationarity for $\boldsymbol{z}(n-k)$. Assuming the vectors $\boldsymbol{z}(n-k)$, $k = 0, \ldots, p-1$, to be distributed according to the same zero-mean Gaussian distribution with covariance matrix $\boldsymbol{\Sigma}_{i1}$, matrix $\boldsymbol{A} = \sum_{k=0}^{p-1} \boldsymbol{z}(n-k)\boldsymbol{z}^\top(n-k)$ is distributed according to a Wishart distribution $\mathcal{W}_2(p, \boldsymbol{\Sigma}_{i1})$ with $p$ degrees of freedom [22, Th. 3.2.4, p. 91]. Thus, $\boldsymbol{d}(n) = \text{diag}(\boldsymbol{A})$ is distributed according to a multivariate gamma distribution with shape parameter $q = p/2$ and $P$ given by (31).

### C. Probability of Error

To simplify the notation, define $t_0$ and $t_1$ such that $\boldsymbol{d}(n) = [\|\boldsymbol{z}_0(n)\|^2, \|\boldsymbol{z}_1(n)\|^2]^\top = [t_0, t_1]^\top$. Also consider $f_j$ to be the bivariate gamma density associated with hypothesis $\mathcal{H}_j$. Then, the probability of error $P_{ij} = P(\mathcal{H}_i|\mathcal{H}_j)$, can be computed as:

$$P_{ij} = \iint_{\mathcal{D}_i} f_j(t_0, t_1)\, dt_0 dt_1 \quad (25)$$

where $\mathcal{D}_i$ represents the integration limits associated with $\mathcal{H}_i$. A detailed expansion of (25) for all classes is presented in the supplementary document, also available in [23]. The integral (25) was implemented using MATLAB function *integral2.m*.

Figs. 3–5 show the probabilities $P(\mathcal{H}_i|\mathcal{H}_j)$ computed using (25) as functions of $c_x^2 \in [0, 10]$ for different sets of parameters. Each row of these figures corresponds to a given true hypothesis $\mathcal{H}_i$, $i = 0, \ldots, 3$. Fig. 3 shows $P(\mathcal{H}_i|\mathcal{H}_j)$ for $\sigma_1^2 = 1$, $\sigma_0^2 = 0.001$, and $p \in \{1, 4, 8, 16, 32\}$. These plots clearly show that the performance of the classifier improves by increasing $c_x^2$ or $p$. A large value of $p$ is especially important in distinguishing between hypotheses $\mathcal{H}_2$ and $\mathcal{H}_3$. It is also clear that the classification error increases significantly for low values of $c_x^2$. As a limiting situation, the vector $\boldsymbol{d}(n)$ will be placed exactly on the line $\|\boldsymbol{z}_0(n)\|^2 = \|\boldsymbol{z}_1(n)\|^2$ separating the classes $\mathcal{H}_0$ and $\mathcal{H}_1$, or $\mathcal{H}_2$ and $\mathcal{H}_3$ (see Fig. 2) for $c_x^2 = 0$.

Since $p = 32$ yielded good classification performance, we opted for fixing $p = 32$ in Figs. 4 and 5, while varying the DT power in Fig. 4 and the noise power in Fig. 5. Although the DT power has little influence on the classifier performance under $\mathcal{H}_0$ and $\mathcal{H}_1$ hypotheses (Fig. 4), a clearer influence is observable under $\mathcal{H}_2$ and $\mathcal{H}_3$. In this case, increasing the DT power tends to increase $P(\mathcal{H}_2|\mathcal{H}_3)$ and $P(\mathcal{H}_3|\mathcal{H}_2)$ (bottom two rows of Fig. 4). This behavior is expected as the effect of a channel change in distinguishing between hypotheses $\mathcal{H}_2$ and $\mathcal{H}_3$ diminishes with the increase of DT power. Fig. 5 explores the effect of the noise power on the classifier performance. It can be noted that a large noise power increases the probability of error in detecting the onset of DT ($P(\mathcal{H}_2|\mathcal{H}_0)$, $P(\mathcal{H}_3|\mathcal{H}_1)$, $P(\mathcal{H}_0|\mathcal{H}_2)$, $P(\mathcal{H}_1|\mathcal{H}_3)$), as the performance is a function of the DT to noise ratio $\sigma_1^2/\sigma_0^2$. This effect, however, is very small for ratios larger than 3 dB, which is typical in practice. Simulations for the one-sample case with different DT and noise powers are available in the supplementary document of this paper. Although the results obtained for the one-sample case show (as expected) a stronger influence of DT and noise power in the classification performance when compared to the results for $p = 32$, they corroborate the above conclusions.

## IV. Monte Carlo Simulations

In this section Monte Carlo (MC) simulations are performed and compared with the theoretical expressions derived in the previous section. These results are also valuable to assess the effect of the independence approximation on the analysis accuracy.

To generate the statistics $\boldsymbol{d}(n)$ by sampling the $(2p)$-dimensional vectors $\boldsymbol{z}_{2d}(n)$ from $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{i2})$, we need to define the covariance ($\boldsymbol{\Sigma}_x$) and correlation ($\boldsymbol{R}_{kx}$) matrices. Considering the input signal to be auto-regressive of order 1 (AR-1), $\boldsymbol{\Sigma}_x$ was chosen as follows [19]:

$$\boldsymbol{\Sigma}_x = \sigma_x^2 \begin{pmatrix} 1 & \rho & \cdots & \rho^{N-1} \\ \rho & 1 & \cdots & \rho^{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \cdots & 1 \end{pmatrix} \quad (26)$$

where $\rho$ controls the input signal correlation. Thus, the entries of $\boldsymbol{R}_{kx} = E\left[\boldsymbol{x}(n)\boldsymbol{x}^\top(n-k)\right]$ can be written as

$$[\boldsymbol{R}_{kx}]_{ij} = \sigma_x^2 \rho^{|i-j-k|}. \quad (27)$$

Note that by fixing the vectors $\boldsymbol{h}_0$ and $\boldsymbol{h}_1$, $\boldsymbol{H}_x$ depends only on $\sigma_x^2$, and $\rho$. Thus, for a given $c_x^2$, $\sigma_x^2$ can be easily computed

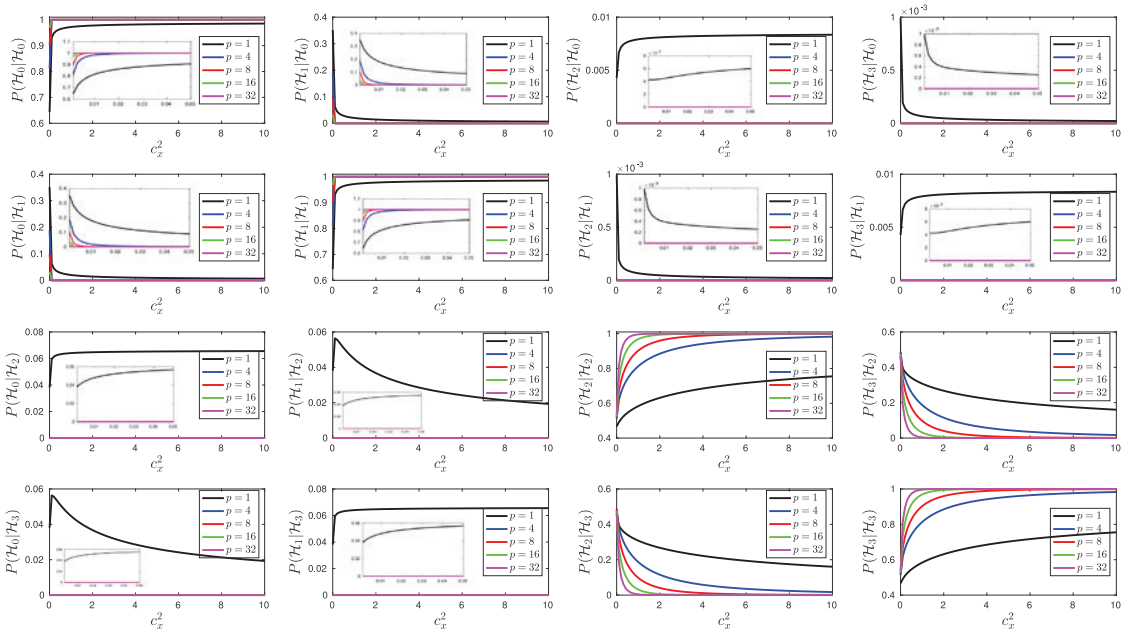

Fig. 3. Theoretical performance curves for single- and multi-sample cases ($\sigma_1^2 = 1, \sigma_0^2 = 0.001$).

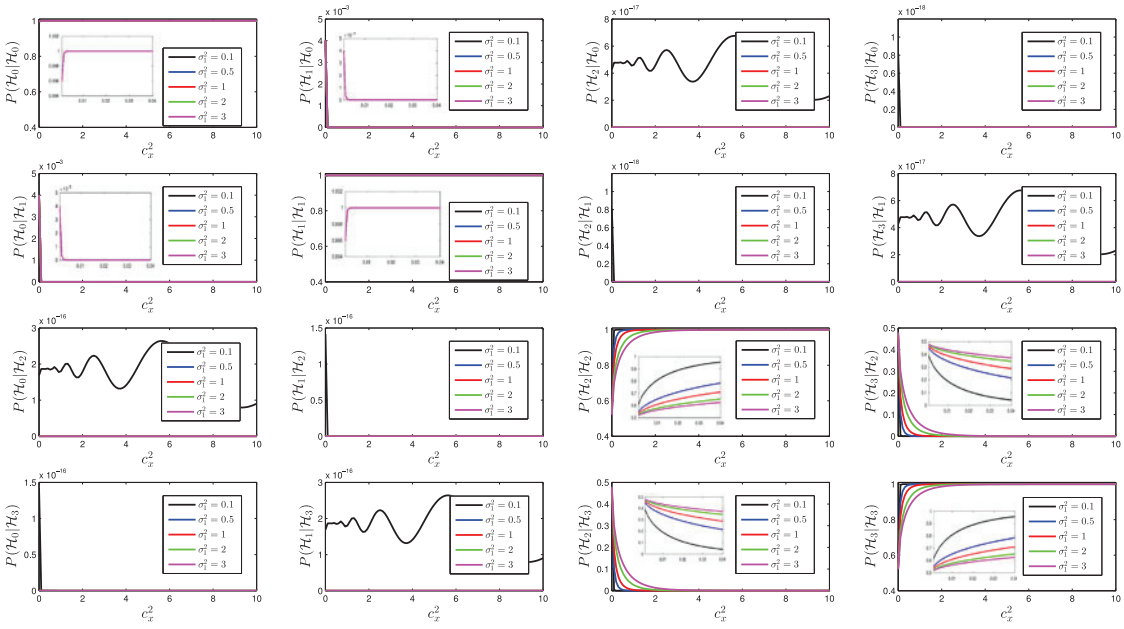

Fig. 4. Theoretical performance curves for different values of DT power ($p = 32$, $\sigma_0^2 = 0.001$).

using (26) and (7). The vectors $\boldsymbol{h}_0$ and $\boldsymbol{h}_1$ were assumed to have 1024 samples, and were constructed using the one-sided exponential channels (see [19] and [20])

$$\boldsymbol{h}_i(k) = \begin{cases} c(0.95)^{k-\Delta_i}, & k \geq \Delta_i \\ 0, & \text{otherwise} \end{cases} \tag{28}$$

where $\Delta_i$ is a relative delay of the channel $\boldsymbol{h}_i$ and the parameter $c$ is defined by the filter gain $G = \boldsymbol{h}_0^\top \boldsymbol{h}_0 = \boldsymbol{h}_1^\top \boldsymbol{h}_1$. Two different scenarios are studied here corresponding to $G = -10$ dB (electrical application) and $G = 6$ dB (acoustic application).[3]

Fig. 6 presents the MC simulations obtained by averaging $10^6$ runs for $G = -10$ dB, with $c_x^2$ varied in the range $[0, 10]$, $\rho = 0.5$, $\sigma_1^2 = 1$, and $\sigma_0^2 = 0.001$, leading to an SNR of 30 dB. When comparing Fig. 6 with theoretical results (Fig. 3), only a very small degradation in classification accuracy is noted, mainly for $\mathcal{H}_2$ and $\mathcal{H}_3$, and $p > 1$. This small difference is attributed to the use of the independence approximation.

MC simulations for different values of the correlation coefficient $\rho$ are available in the supplementary document. Although varying $\rho$ has little impact on the classification performance, it

[3]Since the performance of the MC simulations using $G = 6$ dB are in agreement with the simulations using $G = -10$ dB we suppressed their results from this manuscript. However, the interested reader can find them in the supplementary document.

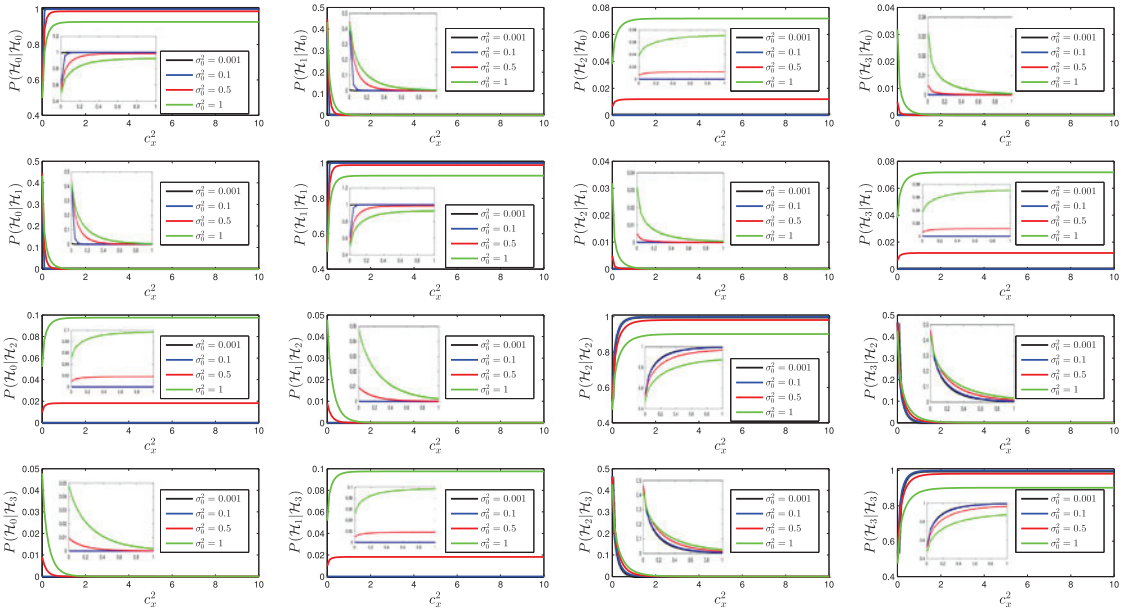

Fig. 5. Theoretical performance curves for different values of noise power ($p = 32$, $\sigma_1^2 = 1$).

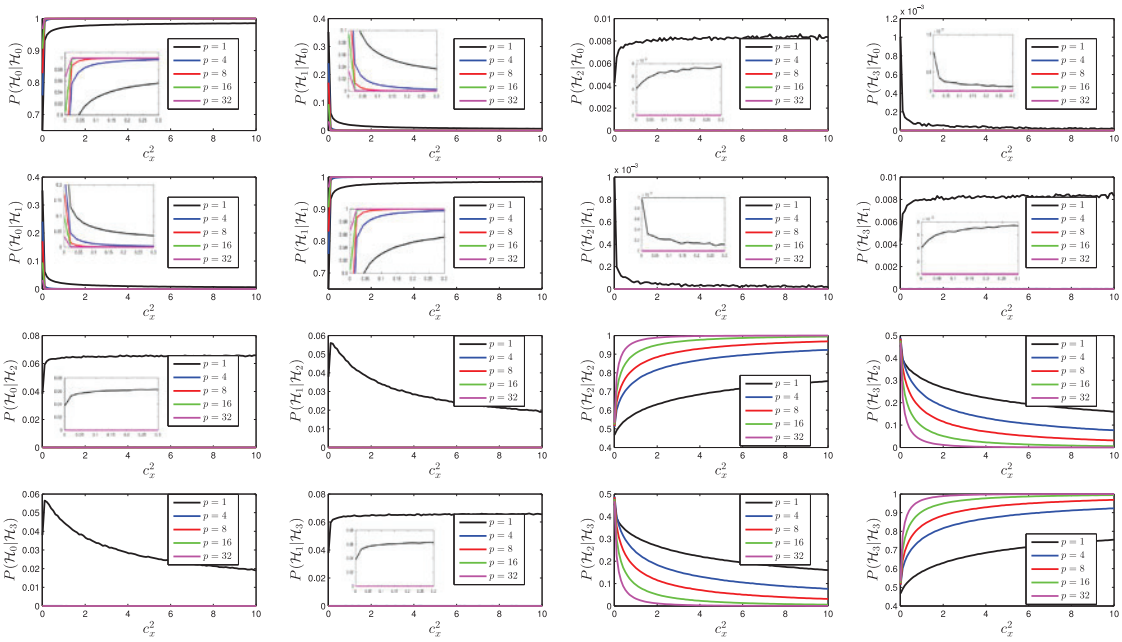

Fig. 6. MC performance curves assuming AR-1 input signal, $\boldsymbol{z}_{pd}(n)$ sampled from $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{ip})$, $G = -10$ dB (electric application), $\sigma_1^2 = 1$, $\sigma_0^2 = 0.001$, $\rho = 0.5$.

is interesting to notice that increasing $\rho$ slightly improves the classification performance in all classes, but especially for $\mathcal{H}_2$ and $\mathcal{H}_3$. This behavior is expected since, for a given $\sigma_x^2$, increasing the correlation of the far-end signal tends to emphasize the effect of the difference $(\boldsymbol{h}_1 - \boldsymbol{h}_0)$ on the values of $\|\boldsymbol{z}_1(n)\|^2$ and $\|\boldsymbol{z}_2(n)\|^2$, facilitating detection of hypotheses in (22).

## V. Application to Echo Cancellers

### A. Control Strategy

The classification hypotheses presented in (2) considered that in each case the adaptive filter had time to converge or diverge. This becomes a critical point for designing the control block (see, Fig. 1) since the probabilities of error are high for low values of $c_x^2$. Two direct consequences related to this characteristic are the following:

1) $(\mathcal{H}_0/\mathcal{H}_1)$ : Whenever $\boldsymbol{h}_0$ is copied to $\boldsymbol{h}_1$ $c_x^2$ becomes zero and the probability of error becomes large between classes $\mathcal{H}_0$ and $\mathcal{H}_1$. In fact, if $\boldsymbol{h}_0 = \boldsymbol{h}_1$ the vector $\boldsymbol{d}(n)$ will be exactly in the frontier between the two classes (see, Fig. 2).
2) $(\mathcal{H}_1/\mathcal{H}_2, \mathcal{H}_3)$ : When CC happens, $\boldsymbol{h}_0$ and $\boldsymbol{h}_1$ may assume values very far from the new true filter response $\boldsymbol{h}_{\text{new}}$. If this is the case, classification errors $(\mathcal{H}_2|\mathcal{H}_1$ or

$\mathcal{H}_3|\mathcal{H}_1$) are expected since both norms $\|\boldsymbol{z}_i(n)\|^2, i = 1, 2$, may become larger than $T$.

To address these problems, we propose a control strategy that combines tuning of the adaptive stepsize $\mu$, defining an appropriate frequency for the realization of the tests, and introducing a delay before actually changing the system state after each decision.

**Adaptation step**

The shadow filter $\boldsymbol{h}_0$ is always adapting, even during DT, since the difference between $\boldsymbol{h}_0$ and $\boldsymbol{h}_1$ is crucial for improving classification rates. However, different adaptation stepsizes can be adopted for each class:

1) During $\mathcal{H}_0$, $\mu = \mu_0$ should be low since the aim is to make a fine tuning of the filter coefficients.
2) During $\mathcal{H}_1$, $\mu = \mu_1$ should be set as high as possible to speed-up convergence of the adaptive algorithm.
3) During $\mathcal{H}_2$, $\mu = \mu_2$ should be set to a small value so that $\boldsymbol{h}_0$ can diverge slowly under DT, start to converge once DT is over or in the occurrence of CC.
4) Class $\mathcal{H}_3$ is critical since it corresponds to the occurrence of CC with or without DT signal. Our practical experience indicates that setting $\mu = \mu_3$ to a value between $\mu_0$ and $\mu_1$ leads to good classification results.

**Frequency of tests**

The difference filter $\boldsymbol{h}_0 - \boldsymbol{h}_1$ plays a central role in classification accuracy. Hence, it is advisable to allow a minimum number $N_t$ of samples between two tests to allow a clear differentiation of the two responses.

**Filter copy**

Whenever classes $\mathcal{H}_0$ or $\mathcal{H}_1$ are detected, the shadow filter $\boldsymbol{h}_0$ should be copied to $\boldsymbol{h}_1$ if $\|\boldsymbol{z}_0(n)\|^2 < \|\boldsymbol{z}_1(n)\|^2$. To account for transients occurring after the exit of a given state (especially when DT stops), it is advisable to consider a delay of $N_c < N_t$ samples between the decision moment and the actual filter copy.

**Decisions in the neighborhood of $\|\boldsymbol{z}_0(n)\|^2 = \|\boldsymbol{z}_1(n)\|^2$**

Decision between $\mathcal{H}_0$ and $\mathcal{H}_1$, and between $\mathcal{H}_2$ and $\mathcal{H}_3$ are rather arbitrary in practical situations when $\|\boldsymbol{z}_0(n)\|^2 \approx \|\boldsymbol{z}_1(n)\|^2$. To address this issue, we propose to allow changes between classes $\mathcal{H}_0$ and $\mathcal{H}_1$, or between $\mathcal{H}_2$ and $\mathcal{H}_3$ only if

$$1 - \varepsilon \leq \frac{\|\boldsymbol{z}_0(n)\|^2}{\|\boldsymbol{z}_1(n)\|^2} \leq 1 + \varepsilon$$

where $\varepsilon \in [0, 1)$.

### B. Synthetic Data

This section considers the AR-1 ($\rho = 0.5$) data discussed in Section IV, and also used in [19], [20]. We considered filter responses $\boldsymbol{h}_0$ and $\boldsymbol{h}_1$ with $N = 1024$ samples, and fixed the parameters $p = 32$, $\sigma_0^2 = 0.001$, and $\sigma_1^2 = 1$. The signal $y(n)$ consisting of 140 K samples (K = 1000) was formulated as

$$y(n) = \begin{cases} \boldsymbol{g}_0^\top \boldsymbol{x}(n) + n_0(n), & n \in \mathcal{I}_1 \\ \boldsymbol{g}_1^\top \boldsymbol{x}(n) + n_0(n), & n \in \mathcal{I}_2 \\ \boldsymbol{g}_1^\top \boldsymbol{x}(n) + n_1(n) + n_0(n), & n \in \mathcal{I}_3 \\ \boldsymbol{g}_2^\top \boldsymbol{x}(n) + n_1(n) + n_0(n), & n \in \mathcal{I}_4 \\ \boldsymbol{g}_2^\top \boldsymbol{x}(n) + n_0(n), & n \in \mathcal{I}_5 \end{cases} \quad (29)$$

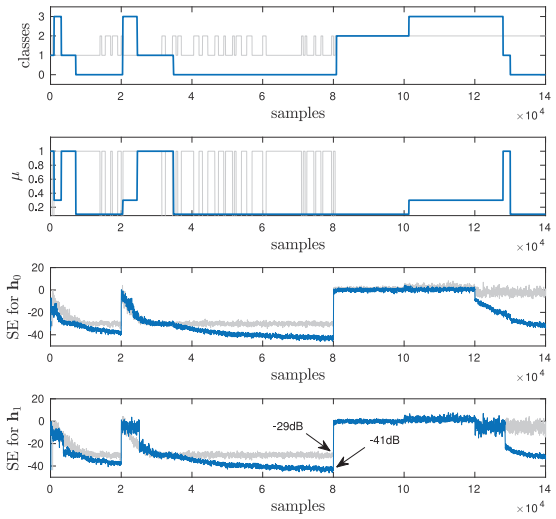


Fig. 7. Performance of the echo canceller system ($G = -10$ dB). From top down, the panels present the evolution of the classification result (top), adaptation stepsize $\mu$, SE in dB for $\boldsymbol{h}_0$ and $\boldsymbol{h}_1$ (bottom). $\sigma_1^2 = 1$, $\sigma_0^2 = 0.001$, $\varepsilon = 0.25$, $N_t = 1024$, $N_c = 512$.

with intervals $\mathcal{I}_1 = [0, 20\,\text{K}]$, $\mathcal{I}_2 = (20\,\text{K}, 80\,\text{K}]$, $\mathcal{I}_3 = (80\,\text{K}, 100\,\text{K}]$, $\mathcal{I}_4 = (100\,\text{K}, 120\,\text{K}]$, and $\mathcal{I}_5 = (120\,\text{K}, 140\,\text{K}]$, and $\boldsymbol{g}_i$, $i \in \{1, 2, 3\}$, being different echo path responses. Hence, CC occurs at sample 20,001, DT occurs between samples 80,001 and 120,000, and a second CC occurs during the DT period at sample 100,001. For comparison, we consider also the GLRT-based strategy proposed in [20]. The adaptive algorithm employed for both methods was the Normalized Least Mean Square (NLMS) algorithm, whose maximum convergence speed is known to be attained for $\mu = 1$ [24]. The control parameters for the proposed strategy were set to $N_t = 1024$, $N_c = 512$, $\mu_0 = \mu_2 = 0.1$, $\mu_1 = 1$, $\mu_3 = 0.3$, and $\varepsilon$ was set to 0.25 for $G = -10$ dB. The GLRT parameters were set following recommendations in [20], with $p = 500$, and the detection threshold $\gamma$ selected to avoid filter copies during DT. For both strategies the adaptive filter coefficients were initialized equal to zero and the adaptation step was initialized as $\mu = \mu_1$ (CC). The simulation results for one realization of the synthetic signal are shown in Fig. 7, where blue curves correspond to the proposed method and gray curves to the GLRT. The top panel presents the classes attributed by the classifier to each sample in time. The second panel presents the step-size corresponding to each class. The squared excess errors (SE) $e_i^2(n) = (y(n) - \boldsymbol{h}_i^\top \boldsymbol{x}(n) - n_0(n))^2$, $i = \{0, 1\}$, for $\boldsymbol{h}_0$ and $\boldsymbol{h}_1$ follow in the bottom two panels. Although the good classification performance is evident in this example, the $\mathcal{H}_1/\mathcal{H}_2, \mathcal{H}_3$ issue discussed in Section V-A can be noticed after the CC at sample 20 K. The samples are classified as $\mathcal{H}_3$ before $\|\boldsymbol{z}_0(n)\|^2$ becomes smaller then $T_p$. Then the correct class $\mathcal{H}_1$ is selected before sample 30 K. However, since the adaptive filter never stops adapting, this problem is satisfactorily mitigated without severe deterioration of the echo canceler performance, as can be verified by the SE results in the two bottom panels. These results clearly show the performance improvement resulting from the generalization of the approach proposed in [19], [20]. The improvement shows especially

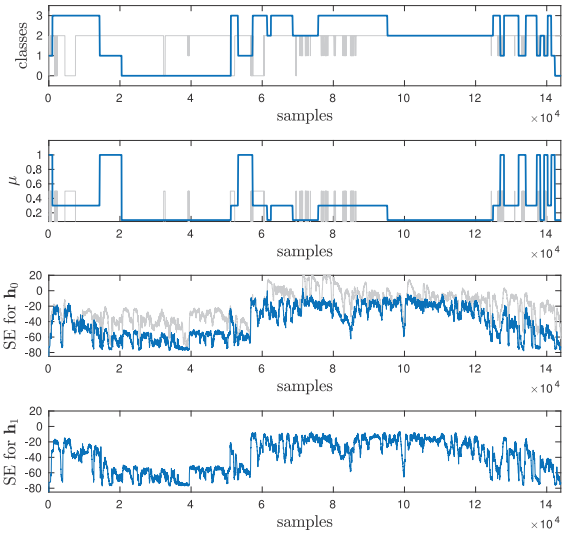


Fig. 8. Performance of the echo canceller system for voice over real channels. From top down, the panels present the evolution of the classification result (top), adaptation stepsize $\mu$, SE in dB for $\boldsymbol{h}_0$ and $\boldsymbol{h}_1$ (bottom). $\varepsilon = 0.25$, $N_t = 1024$, $N_c = 512$. Results for the proposed method (blue) and using the method in [18] (gray).

during the single-talk periods. As DT or CC do not occur during these periods, the proposed solution leads to a reduction of the stepsize $\mu$, clearly improving the quality of channel estimation. Note, for instance, that the stepsize reduction that happens at iteration 35 K due to the acceptance of hypothesis $\mathcal{H}_0$ leads to a drop in SE that reaches 12 dB at iteration 80 K. A decision threshold could not be found for the GLRT method that avoided CC classification during DT and at the same time allowed accurate classification after sample 120 K. Any threshold leading to the correct classification of this portion of the signal also led to $\mathcal{H}_1$ classification during the $\mathcal{H}_3$ periods in interval $\mathcal{I}_4$. This also shows the benefits of modeling this extra hypothesis.

Simulations with $G = 6$ dB yielded similar results, and are available in the supplementary document [23].

### C. Voice Data Over a Real Channel

For the simulation presented in this section we used the same voice data and channels considered in [19], [20]. The data is approximately 144 K samples long, with two CC's occurring at sample 50 K and 123 K, and an intense DT occurring between 57–123 K. The simulation results presented in Fig. 8 compare the proposed decision framework (blue) with the sequential classification strategy presented in [18] (gray). To deal with the power fluctuation inherent in speech signals, we used $p = 500$ and set the detection threshold $T_p = 1 \times 10^{-5}$ chosen empirically to avoid $\mathcal{H}_0$ and $\mathcal{H}_1$ errors during DT. The remaining control strategy parameters were kept the same used in the synthetic simulation presented in Fig. 7. The parameters used for the method in [18] were set to the same values used by the authors. Although the detector presented in [18] also considers different classes, the authors did not consider the influence of multiple samples nor used a shadow filter configuration, which clearly impacts the results. The results displayed in Fig. 8 can be also compared with the result obtained in [20, Fig. 9], which indicates that the proposed classification and control strategies perform at least as well as previous echo cancellation systems.

## VI. Results and Conclusions

In this manuscript we presented a low computational cost multi-class classifier with a coupling control strategy for the echo cancellation problem. The proposed classification rule initially proposed for one-sample was easily extended to the multi-sample scenario. Error probabilities were also analytically computed under the assumption of independence among vectors $\boldsymbol{z}(n-k)$. This assumption led to bivariate gamma distributions for the sufficient statistics $\boldsymbol{d}(n)$ and performance curves that proved accurate when confronted with Monte Carlo Simulations. The results showed that the greater flexibility provided by the multi-class approach could be well explored by the control strategy which considered different step-sizes under each hypothesis. The simulations with synthetic data showed that the multi-class strategy is viable if accurate double-talk and noise power can be estimated, improving the filter convergence during long periods of single-talk. Simulations in a more realistic scenario (voice over real channels) showed that the proposed strategy works as well as other methods in the literature even ignoring the power fluctuation of speech signals and using a fixed threshold $T_p$.

## Appendix A
## Classification Rule

This appendix derives the classification rule (10) for the one sample case. This rule corresponds to accepting hypothesis $\mathcal{H}_i$ if

$$\boldsymbol{z}^\top(n)\left(\boldsymbol{\Sigma}_{j1}^{-1} - \boldsymbol{\Sigma}_{i1}^{-1}\right)\boldsymbol{z}(n) > \ln\left(\frac{|\boldsymbol{\Sigma}_{i1}|}{|\boldsymbol{\Sigma}_{j1}|}\right) \tag{30}$$

for all $j \neq i$. As a consequence hypothesis $\mathcal{H}_0$ is accepted if the three following conditions are satisfied

$$\boldsymbol{z}^\top(n)\left(\boldsymbol{\Sigma}_{11}^{-1} - \boldsymbol{\Sigma}_{01}^{-1}\right)\boldsymbol{z}(n) > \ln\left(\frac{|\boldsymbol{\Sigma}_{01}|}{|\boldsymbol{\Sigma}_{11}|}\right)$$

$$\boldsymbol{z}^\top(n)\left(\boldsymbol{\Sigma}_{21}^{-1} - \boldsymbol{\Sigma}_{01}^{-1}\right)\boldsymbol{z}(n) > \ln\left(\frac{|\boldsymbol{\Sigma}_{01}|}{|\boldsymbol{\Sigma}_{21}|}\right)$$

$$\boldsymbol{z}^\top(n)\left(\boldsymbol{\Sigma}_{31}^{-1} - \boldsymbol{\Sigma}_{01}^{-1}\right)\boldsymbol{z}(n) > \ln\left(\frac{|\boldsymbol{\Sigma}_{01}|}{|\boldsymbol{\Sigma}_{31}|}\right).$$

By replacing the matrix inverses and determinants in these expressions, the following results are obtained

$$z_0^2(n) - z_1^2(n) > 0$$

$$z_1^2(n)\left(\frac{1}{\sigma_0^2+\sigma_1^2} - \frac{1}{\sigma_0^2}\right) > \ln\left(1+\frac{\sigma_1^2}{\sigma_0^2}\right)$$

$$\frac{z_0^2(n)}{\sigma_0^2+\sigma_1^2} - \frac{z_1^2(n)}{\sigma_0^2} > -\ln\left(1+\frac{\sigma_1^2}{\sigma_0^2}\right).$$

These three conditions are equivalent to

$$z_1^2(n) < z_0^2(n) \text{ and } z_1^2(n) < \frac{\sigma_0^2(\sigma_0^2+\sigma_1^2)}{\sigma_1^2}\ln\left(1+\frac{\sigma_1^2}{\sigma_0^2}\right).$$

Hypothesis $\mathcal{H}_1$ is accepted if the three following conditions are satisfied

$$\boldsymbol{z}^\top(n)\left(\boldsymbol{\Sigma}_{01}^{-1}-\boldsymbol{\Sigma}_{11}^{-1}\right)\boldsymbol{z}(n)>\ln\left(\frac{|\boldsymbol{\Sigma}_{11}|}{|\boldsymbol{\Sigma}_{01}|}\right)$$
$$\boldsymbol{z}^\top(n)\left(\boldsymbol{\Sigma}_{21}^{-1}-\boldsymbol{\Sigma}_{11}^{-1}\right)\boldsymbol{z}(n)>\ln\left(\frac{|\boldsymbol{\Sigma}_{11}|}{|\boldsymbol{\Sigma}_{21}|}\right)$$
$$\boldsymbol{z}^\top(n)\left(\boldsymbol{\Sigma}_{31}^{-1}-\boldsymbol{\Sigma}_{11}^{-1}\right)\boldsymbol{z}(n)>\ln\left(\frac{|\boldsymbol{\Sigma}_{11}|}{|\boldsymbol{\Sigma}_{31}|}\right).$$

Equivalently

$$z_1^2(n)-z_0^2(n)>0$$
$$\frac{z_1^2(n)}{\sigma_0^2+\sigma_1^2}-\frac{z_0^2(n)}{\sigma_0^2}>-\ln\left(1+\frac{\sigma_1^2}{\sigma_0^2}\right)$$
$$z_0^2(n)\left(\frac{1}{\sigma_0^2+\sigma_1^2}-\frac{1}{\sigma_0^2}\right)>-\ln\left(1+\frac{\sigma_1^2}{\sigma_0^2}\right).$$

These three conditions are equivalent to

$$z_1^2(n)>z_0^2(n)\text{ and }z_0^2(n)<\frac{\sigma_0^2(\sigma_0^2+\sigma_1^2)}{\sigma_1^2}\ln\left(1+\frac{\sigma_1^2}{\sigma_0^2}\right).$$

Hypothesis $\mathcal{H}_2$ is accepted if the three following conditions are satisfied

$$\boldsymbol{z}^\top(n)\left(\boldsymbol{\Sigma}_{01}^{-1}-\boldsymbol{\Sigma}_{21}^{-1}\right)\boldsymbol{z}(n)>\ln\left(\frac{|\boldsymbol{\Sigma}_{21}|}{|\boldsymbol{\Sigma}_{01}|}\right)$$
$$\boldsymbol{z}^\top(n)\left(\boldsymbol{\Sigma}_{11}^{-1}-\boldsymbol{\Sigma}_{21}^{-1}\right)\boldsymbol{z}(n)>\ln\left(\frac{|\boldsymbol{\Sigma}_{21}|}{|\boldsymbol{\Sigma}_{11}|}\right)$$
$$\boldsymbol{z}^\top(n)\left(\boldsymbol{\Sigma}_{31}^{-1}-\boldsymbol{\Sigma}_{21}^{-1}\right)\boldsymbol{z}(n)>\ln\left(\frac{|\boldsymbol{\Sigma}_{21}|}{|\boldsymbol{\Sigma}_{31}|}\right).$$

Equivalently

$$z_1^2(n)>\frac{\sigma_0^2(\sigma_0^2+\sigma_1^2)}{\sigma_1^2}\ln\left(1+\frac{\sigma_1^2}{\sigma_0^2}\right)$$
$$\frac{z_1^2(n)}{\sigma_0^2+\sigma_1^2}-\frac{z_0^2(n)}{\sigma_0^2}<-\ln\left(1+\frac{\sigma_1^2}{\sigma_0^2}\right)$$
$$z_0^2(n)>z_1^2(n).$$

These three conditions are equivalent to

$$z_1^2(n)<z_0^2(n)\text{ and }z_1^2(n)>\frac{\sigma_0^2(\sigma_0^2+\sigma_1^2)}{\sigma_1^2}\ln\left(1+\frac{\sigma_1^2}{\sigma_0^2}\right).$$

Hypothesis $\mathcal{H}_3$ is accepted if the three following conditions are satisfied

$$\boldsymbol{z}^\top(n)\left(\boldsymbol{\Sigma}_{01}^{-1}-\boldsymbol{\Sigma}_{31}^{-1}\right)\boldsymbol{z}(n)>\ln\left(\frac{|\boldsymbol{\Sigma}_{31}|}{|\boldsymbol{\Sigma}_{01}|}\right)$$
$$\boldsymbol{z}^\top(n)\left(\boldsymbol{\Sigma}_{11}^{-1}-\boldsymbol{\Sigma}_{31}^{-1}\right)\boldsymbol{z}(n)>\ln\left(\frac{|\boldsymbol{\Sigma}_{31}|}{|\boldsymbol{\Sigma}_{11}|}\right)$$
$$\boldsymbol{z}^\top(n)\left(\boldsymbol{\Sigma}_{21}^{-1}-\boldsymbol{\Sigma}_{31}^{-1}\right)\boldsymbol{z}(n)>\ln\left(\frac{|\boldsymbol{\Sigma}_{31}|}{|\boldsymbol{\Sigma}_{21}|}\right).$$

Equivalently

$$z_1^2(n)-\frac{z_0^2(n)}{\sigma_0^2+\sigma_1^2}>\ln\left(1+\frac{\sigma_1^2}{\sigma_0^2}\right)$$
$$z_0^2(n)>\frac{\sigma_0^2(\sigma_0^2+\sigma_1^2)}{\sigma_1^2}\ln\left(1+\frac{\sigma_1^2}{\sigma_0^2}\right)$$
$$z_0^2(n)<z_1^2(n).$$

These three conditions are equivalent to

$$z_1^2(n)>z_0^2(n)\text{ and }z_0^2(n)>\frac{\sigma_0^2(\sigma_0^2+\sigma_1^2)}{\sigma_1^2}\ln\left(1+\frac{\sigma_1^2}{\sigma_0^2}\right).$$

## Appendix B
## Multivariate Gamma Distribution

Define $p$ independent random vectors of $\mathbb{R}^2$ denoted as $\boldsymbol{v}_k(\ell)=[v_0(\ell-k),v_1(\ell-k)]^\top\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{\Sigma}),\quad k=0,\ldots,p-1$, and the $2\times p$ matrix $\boldsymbol{V}(\ell)=[\boldsymbol{v}_0(\ell),\boldsymbol{v}_1(\ell),\ldots,\boldsymbol{v}_{p-1}(\ell)]$. Then, the $2\times 2$ matrix $\boldsymbol{A}=\boldsymbol{V}(\ell)\boldsymbol{V}^\top(\ell)$ is known to be distributed according to a Wishart distribution $\mathcal{W}_2(p,\boldsymbol{\Sigma})$ with $p$ degrees of freedom and covariance matrix $\boldsymbol{\Sigma}$ [22, Th. 3.2.4, p. 91]. Now, define the vector $\boldsymbol{d}$ composed by the elements of the main diagonal of $\boldsymbol{A}$. Then, it was shown in Proposition 1.3.3 in [25, p. 32] that $\boldsymbol{d}$ is distributed according to a multivariate gamma distribution denoted $\mathcal{G}(q,P)$ with shape parameter $q=p/2$ and scale parameter $P=\{p_1,p_2,p_{12}\}$, with

$$\begin{aligned}p_1&=2\boldsymbol{\Sigma}(1,1)\\p_2&=2\boldsymbol{\Sigma}(2,2)\\p_{12}&=4\left[\boldsymbol{\Sigma}(1,1)\boldsymbol{\Sigma}(2,2)-\boldsymbol{\Sigma}(1,2)\boldsymbol{\Sigma}(2,1)\right]\end{aligned}\qquad(31)$$

where $\boldsymbol{\Sigma}(1,1)$, $\boldsymbol{\Sigma}(1,2)=\boldsymbol{\Sigma}(2,1)$ and $\boldsymbol{\Sigma}(2,2)$ are the elements of the covariance matrix $\boldsymbol{\Sigma}$.

Now, making $\boldsymbol{v}_k(n)=\boldsymbol{z}(n-k)=[z_0(n-k),z_1(n-k)]^\top\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{\Sigma}_{ip}),\quad k=0,\ldots,p-1$, for each hypothesis $\mathcal{H}_i$, and assuming the independence of $\boldsymbol{z}(n-i)$ and $\boldsymbol{z}(n-j)$ for $i\neq j$[4], the above results show that $\boldsymbol{d}(n)=[\|\boldsymbol{z}_0(n)\|^2,\|\boldsymbol{z}_1(n)\|^2]^\top$ is distributed according to a multivariate gamma distribution with shape parameter $q=p/2$ and scale parameter $P=\{p_1,p_2,p_{12}\}$ evaluated from (31) with $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_{ip}$.

## References


[1] J. Gunther and T. Moon, "Adaptive cancellation of acoustic echoes during double-talk based on an information theoretic criteria," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2009, pp. 650–654.

[2] J. Gunther and T. Moon, "Blind acoustic echo cancellation without double-talk detection," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, New Paltz, NY, USA, Oct. 2015, pp. 1–5.

[3] A. Mader, H. Puder, and G. Schmidt, "Step-size control for echo cancellation filters—An overview," *Signal Process.*, vol. 80, no. 9, pp. 1697–1719, Sep. 2000.

[4] 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, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168414001157


[4]This is a simplifying assumption employed for mathematical tractability. Simulation results will show that this assumption has little impact on the classifier performance.

[5] F. Yang, G. Enzner, and J. Yang, "Statistical convergence analysis for optimal control of DFT-domain adaptive echo canceler," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 5, pp. 1095–1106, May 2017.
[6] G. Rombouts, T. van Waterschoot, K. Struyve, and M. Moonen, "Acoustic feedback cancellation for long acoustic paths using a nonstationary source model," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3426–3434, Sep. 2006.
[7] 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, no. 3, pp. 846–858, Mar. 2007.
[8] J. J. Shynk *et al.*, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Process. Mag.*, vol. 9, no. 1, pp. 14–37, Jan. 1992.
[9] K. Ochiai, T. Araseki, and T. Ogihara, "Echo canceller with two echo path models," *IEEE Trans. Commun.*, vol. 25, no. 6, pp. 589–595, Jun. 1977.
[10] J. Benesty, D. R. Morgan, and J. H. Cho, "A new class of doubletalk detectors based on cross-correlation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 8, no. 2, pp. 168–172, Mar. 2000.
[11] C. Schuldt, F. Lindstrom, and I. Claesson, "A delay-based double-talk detector," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 6, pp. 1725–1733, Aug. 2012.
[12] M. Z. Ikram, "Double-talk detection in acoustic echo cancellers using zero-crossings rate," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Brisbane, QLD, Australia, Apr. 2015, pp. 1121–1125.
[13] P. Åhgren and A. Jakobsson, "A study of doubletalk detection performance in the presence of acoustic echo path changes," *IEEE Trans. Consum. Electron.*, vol. 52, no. 2, pp. 515–522, May 2006.
[14] C. Carlemalm, F. Gustafsson, and B. Wahlberg, "On the problem of detection and discrimination of double talk and change in the echo path," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Atlanta, GA, USA, May 1996, pp. 2742–2745.
[15] C. Carlemalm and A. Logothetis, "On detection of double talk and changes in the echo path using a Markov modulated channel model," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Munich, Germany, Apr. 1997, pp. 3869–3872.
[16] M. Fozunbal, M. C. Hans, and R. W. Schafer, "A decision-making framework for acoustic echo cancellation," in *Proc IEEE Workshop Appl. Signal Process. Audio Acoust.*, New Paltz, NY, USA, Oct. 2005, pp. 33–36.
[17] H. K. Jung, N. S. Kim, and T. Kim, "A new double-talk detector using echo path estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Orlando, FL, USA, May 2002, pp. 1897–1900.
[18] H. K. Jung, N. S. Kim, and T. Kim, "A new double-talk detector using echo path estimation," *Speech Commun.*, vol. 45, no. 1, pp. 41–48, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167639304001037
[19] N. J. Bershad and J.-Y. Tourneret, "Echo cancellation: A likelihood ratio test for double talk versus channel change," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4572–4581, Dec. 2006.
[20] J.-Y. Tourneret, N. J. Bershad, and J. C. M. Bermudez, "Echo cancellation—The generalized likelihood ratio test for double-talk versus channel change," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 916–926, Mar. 2009.
[21] S. M. Kay, *Fundamentals of Statistical Signal Processing. Vol. 1: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
[22] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. Hoboken, NJ, USA: Wiley, 2005.
[23] T. Imbiriba, J. C. M. Bermudez, J.-Y. Tourneret, and N. J. Bershad, "Technical report: A new decision-theory-based framework for echo canceler control," ArXiv e-prints, Nov. 2017.
[24] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1996.
[25] F. Chatelain, "Multivariate gamma based distributions for image processing applications," Ph.D. dissertation, Nat. Polytechn. Inst. Toulouse, Toulouse, France, 2007.

**Tales Imbiriba** (S'14–M'17) received the B.E.E. and M.Sc. degrees from the Federal University of Pará, Belém, Brazil, in 2006 and 2008, respectively, and the Doctorade degree from the Federal University of Santa Catarina (UFSC), Florianópolis, Brazil, in 2016. He is currently a Postdoctoral Researcher with the Digital Signal Processing Laboratory, UFSC. His research interests include audio and image processing, pattern recognition, kernel methods, and adaptive filtering.

**José Carlos M. Bermudez** (S'78–M'85–SM'02) received the B.E.E. degree from the Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil, in 1978, the M.Sc. degree in electrical engineering from COPPE/UFRJ in 1981, and the Ph.D. degree in electrical engineering from Concordia University, Montreal, QC, Canada, in 1985. In 1985, he joined the Department of Electrical Engineering, Federal University of Santa Catarina, Florianópolis, Brazil, where he is currently a Professor. He spent sabbatical years at the University of California, Irvine, CA, USA, in 1994, and at the Institut National Polytechnique de Toulouse, Toulouse, France, in 2012. His recent research interests have been in statistical signal processing, including adaptive filtering, image processing, hyperspectral image processing and machine learning. He was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING in the area of adaptive filtering from 1994 to 1996 and from 1999 to 2001 and an Associate Editor for the *EURASIP Journal of Advances on Signal Processing* from 2006 to 2010. He is a Senior Area Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and an Associated Editor for the *GRETSI Journal Traitement du Signal*. He is a member and Elect Chair of the Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society.

**Jean-Yves Tourneret** (SM'08) received the Ingnieur degree in electrical engineering from the École Nationale Supérieure d'Électronique, d'Électrotechnique, d'Informatique et d'Hydraulique of Toulouse (ENSEEIHT), University of Toulouse, Toulouse, France, in 1989, and the Ph.D. degree from the National Polytechnic Institute of Toulouse, Toulouse, in 1992. He is currently a Professor with ENSEEIHT and a member of the IRIT Laboratory (UMR 5505 of the CNRS). His research activities have centered around statistical signal processing with a particular interest to Markov chain Monte Carlo methods. He was the program chair of EUSIPCO, Toulouse, in 2002. He was also a member of the organizing committee for the 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing, Toulouse. He has been a member of different technical committees including the Signal Processing Theory and Methods Committee of the IEEE Signal Processing Society (from 2001 to 2007 and since 2010). He is an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.

**Neil J. Bershad** (F'88) received the B.E.E. degree from the Rensselaer Polytechnic Institute, Troy, NY, USA, in 1958, the M.S. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 1960, and the Ph.D. degree in electrical engineering from the Rensselaer Polytechnic Institute in 1962. In 1966, he joined the Faculty of the Henry Samueli School of Engineering, University of California, Irvine, CA, where he is currently an Emeritus Professor of electrical engineering and computer science. His research interests have involved stochastic systems modeling and analysis. His recent research interests have been in the area of stochastic analysis of adaptive filters, including the statistical learning behavior of adaptive filter structures for echo cancellation, active acoustic noise cancellation, and variable gain (mu) adaptive algorithms. He was an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS in the area of phase-locked loops and synchronization and for the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING in the area of adaptive filtering.