Performance of ESPRIT for Estimating Mixtures of Complex Exponentials Modulated by Polynomials

Roland Badeau, *Member, IEEE*, Gaël Richard, *Senior Member, IEEE*, and Bertrand David, *Member, IEEE*

***Abstract*—High-resolution (HR) methods are known to provide accurate frequency estimates for discrete spectra. The polynomial amplitude complex exponentials (PACE) model, also called quasi-polynomial model in the literature, was presented as the most general model tractable by HR methods. A subspace-based estimation scheme was recently proposed, derived from the classical ESPRIT algorithm. In this paper, we focus on the performance of this estimator. We first present some asymptotic expansions of the estimated parameters, obtained at the first order under the assumption of a high signal-to-noise ratio (SNR). Then the performance of the generalized ESPRIT algorithm for estimating the parameters of this model is analyzed in terms of bias and variance, and compared to the Cramér-Rao bounds (CRB). This performance is studied in an asymptotic context, and it is proved that the efficiency of undamped single poles estimators is close to the optimality. Moreover, our results show that the best performance is obtained for a proper dimensioning of the data. To illustrate the practical capabilities of the generalized ESPRIT algorithm, we finally propose an application to ARMA filter synthesis, in the context of system conversion from continuous time to discrete time.**

***Index Terms*—ESPRIT, high resolution (HR), multiple eigenvalues, performance analysis, perturbation theory, polynomial modulation.**

I. Introduction

HIGH RESOLUTION (HR) methods, such as the well-known ESPRIT algorithm [1], are very classical techniques for estimating mixtures of complex exponentials in white noise. However, the underlying exponential sinusoidal model (ESM), although the most studied in the literature, is not the most general model tractable by HR methods. Indeed the ESM only accounts for systems with single poles, whereas one can find examples of systems involving multiple poles, which generate mixtures of complex exponentials modulated by polynomials, as shown in [2]. For instance, critically damped harmonic oscillators involve a double pole [3]. Laguerre functions are a special case of signals with multiple poles (the exponentials are modulated by Laguerre polynomials), often used in the estimation of time delays [4], [5], and in biomedical engineering, for modeling fluorescence decay [6]. Signals with multiple poles also appear in quantum physics, as solutions of the Schrödinger equation for hydrogen-like atoms [7], in laser physics, as transverse laser modes [8], and in finance, for modeling the evolution of interest rates [9].

In order to estimate the parameters of this more general polynomial amplitude complex exponentials (PACE) model, also referred to as the quasi-polynomial model [10] in the literature, we proposed in [2] a generalization of the ESPRIT algorithm for estimating multiple poles. The polynomial amplitude parameters can then be recovered by means of a least squares (LS) method. The performance of an estimator is generally described in terms of bias and variance, the latter being generally compared to the Cramér-Rao bound (CRB), in terms of statistical efficiency [11]. An analysis of the CRB for the frequencies and damping factors of complex quasi-polynomials in white noise was proposed in [10]. In [12], we derive analytic expressions of the CRB for the frequencies, damping factors, amplitudes, and phases of quasi-polynomials in colored noise, and these expressions are simplified in an asymptotic context. In particular, it is shown that the CRB for the parameters associated to a multiple pole present an exponential increase with the order of the pole, which suggests that the practical estimation of the PACE model is only possible if the exponentials are modulated by polynomials of low order.

Unfortunately, in the case of HR methods, the bias and variance cannot be calculated analytically, because the extraction of polynomial roots, or matrix eigenvalues, induces a complex relationship between the statistics of the signal and those of the estimators. In the case of the ESM, however, asymptotic results were obtained with the perturbation theory. These results rely either on the hypothesis of a high window length ($N \to +\infty$, in the case where all the poles are on the unit circle), or on the hypothesis of a high signal-to-noise ratio (SNR) ($\mathrm{SNR} \to +\infty$). For instance, it was established in [13] and [14] that the Prony [15] and Pisarenko [16] methods are very inefficient: their variances are much greater than the CRB. Conversely, the minimum norm method [17], MUSIC [18], ESPRIT [1], and Matrix Pencil [19] have an asymptotic efficiency close to 1, as shown in [20]–[24]. More precisely, it was proved in [22] and [23] in the case of undamped sinusoids that MUSIC and ESPRIT perform similarly, with a slight advantage for ESPRIT. This was confirmed in [24] in the more general case of the ESM: ESPRIT and Matrix Pencil are less sensitive to noise than MUSIC.

In the case of the PACE model, it was shown in [2] that the presence of noise scatters the multiple poles into several single poles, forming the vertices of a regular polygon as a first-order approximation. However, the original multiple pole can be recovered by computing the arithmetic mean of the scattered poles. In this paper, we analyze the performance of this approach in presence of *colored* noise and under the high SNR

Manuscript received October 5, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Philippe Loubaton.

The authors are with the Ecole Nationale Supérieure des Télécommunications, Département TSI, 75634 Paris Cedex 13, France (e-mail: roland.badeau@enst.fr; gael.richard@enst.fr; bertrand.david@enst.fr).

Digital Object Identifier 10.1109/TSP.2007.906744

hypothesis, in terms of first-order perturbations. These developments are utilized to show that the estimators proposed in [2] for the PACE model are unbiased, and their variances are calculated and compared to the CRB. Additionally, by considering a high observation length and a white noise, it is shown that the efficiency of the estimators is close to 1. In particular, we generalize a result presented in [19], which provides the ideal dimensioning of the data matrix in order to improve the efficiency of single poles estimators. However, our simulation results confirm that the practical estimation of the parameters is only possible for poles of low order. To illustrate the capabilities of our estimation method, we finally propose an application to ARMA filter synthesis, in the context of system conversion from continuous time to discrete time.

The paper is organized as follows. Section II describes the PACE model and the estimation method introduced in [2]. Then the influence of an additive perturbation onto the estimated frequencies, damping factors, amplitudes and phases is studied in Section III. Section IV analyzes the performance of the estimators: their first-order bias and variances are calculated in Section IV-A, then their asymptotic expansions are derived in Section IV-B (in the case of undamped single poles). These developments are illustrated in Section V-A, where the variances of the estimators are compared to the CRB, and the generalized ESPRIT algorithm is applied to ARMA filter synthesis in Section V-B. The main conclusions of this paper are summarized in Section VI. Finally, the mathematical developments for the perturbation analysis are provided in the Appendix.

II. General Framework

In Sections II-A and II-B, we summarize the basics of the PACE model, also called quasi-polynomial model, and the generalized ESPRIT algorithm, which were presented in [2].

A. Polynomial Amplitude Complex Exponentials

Definition 1: Let $K \in \mathbb{N}^*$. For all $k \in \{0 \dots K-1\}$, define the partial order $M_k \in \mathbb{N}^*$, the frequency $f_k \in]-(1/2),(1/2)]$, the damping (or amplifying) factor $\delta_k \in \mathbb{R}$, and the complex pole $z_k = e^{\delta_k + i2\pi f_k}$. Suppose that the complex poles are distinct from one another. Then a discrete signal $s(t)$ satisfies the PACE model of order $r \triangleq \sum_{k=0}^{K-1} M_k$ if and only if it can be written in the form

$$s(t) = \sum_{k=0}^{K-1} \alpha_k[t] {z_k}^t \tag{1}$$

where $\forall k \in \{0, \dots, K-1\}$, $\alpha_k[t]$ is a complex polynomial of order $M_k - 1$.

The polynomial $\alpha_k[t]$ can be decomposed onto the polynomial basis of falling factorials:

Definition 2 (Falling Factorial): For all $m \in \mathbb{Z}$, the falling factorial of order m is the polynomial[1]

$$F_m[X] = \begin{cases} 0 & \text{if } m < 0 \\ 1 & \text{if } m = 0 \\ \frac{1}{m!} \prod_{m'=0}^{m-1} (X - m') & \text{if } m > 0. \end{cases}$$

In this basis, (1) can be rewritten in the form

$$s(t) = \sum_{k=0}^{K-1} \sum_{m=0}^{M_k-1} \alpha_k^{(m)} F_m[t] {z_k}^{t-m} \tag{2}$$

where $\forall k \in \{0 \dots K-1\}$, $\forall m \in \{0, M_k - 1\}$, $\alpha_k^{(m)}$ is a complex amplitude. Define the real amplitude $a_k^{(m)} = |\alpha_k^{(m)}|$, and the phase[2] $\phi_k^{(m)} = \Im(\ln(\alpha_k^{(m)}))$.

The PACE model can be characterized by means of matrix analysis. Indeed, the samples of the discrete signal $s(t)$ can be arranged into a Hankel data matrix with $n > r$ rows and $l \geq r$ columns

$$\boldsymbol{S} = \begin{bmatrix} s(t-l+1) & \cdots & s(t-1) & s(t) \\ s(t-l+2) & \cdots & s(t) & s(t+1) \\ \vdots & \cdots & \vdots & \vdots \\ s(t-l+n) & \cdots & s(t+n-2) & s(t+n-1) \end{bmatrix}. \tag{3}$$

In particular, the range space of $\boldsymbol{S}$ can be characterized by the *generalized Pascal* and *Pascal-Vandermonde* matrices.

Definition 3 (Generalized Pascal Matrices): Let $z \in \mathbb{C}$ and $M \in \mathbb{N}^*$. The generalized Pascal matrix denoted $\boldsymbol{C}_M^n(z)$ is an $n \times M$ matrix whose coefficients are $\boldsymbol{C}_M^n(z)_{(i,j)} = F_j[i] z^{i-j}$ for all $i \in \{0 \dots n-1\}$ and $j \in \{0 \dots M-1\}$.

Example 4: If $M = 3$ and $n = 6$

$$\boldsymbol{C}_3^6(z) = \begin{bmatrix} 1 & 0 & 0 \\ z & 1 & 0 \\ z^2 & 2z & 1 \\ z^3 & 3z^2 & 3z \\ z^4 & 4z^3 & 6z^2 \\ z^5 & 5z^4 & 10z^3 \end{bmatrix}.$$

Definition 5 (Pascal-Vandermonde Matrices): The $n \times r$ Pascal-Vandermonde matrix is the matrix formed by concatenating the generalized Pascal matrices $\boldsymbol{C}_k^n = \boldsymbol{C}_{M_k}^n(z_k)$

$$\boldsymbol{V}^n = \left[\boldsymbol{C}_0^n, \dots, \boldsymbol{C}_{(K-1)}^n \right].$$

Based on the above definitions, the following proposition, which is proved in [2], shows a factorization of the Hankel data matrix:[3]

Proposition 1 (Factorization of the Data Matrix): An $n \times l$ Hankel matrix $\boldsymbol{S}$ of the form (3) where $s(t)$ is the signal defined in (2) can be factorized in the form

$$\boldsymbol{S} = \boldsymbol{V}^n \boldsymbol{D} {\boldsymbol{V}^l}^T \tag{4}$$

[1]Note that this definition does not match the classical definition of the falling factorial [25], [26], from which the multiplicative factor $1/m!$ is missing.

[2]In the whole paper, the notation $\ln(.)$ denotes the determination of the complex logarithm which corresponds to an angle lying in $]-\pi, \pi[$.

[3]In linear systems realization theory, state space representations also lead to low-rank factorizations of Hankel matrices [27].

where $\boldsymbol{V}^n$ and $\boldsymbol{V}^l$ are the $n \times r$ and $l \times r$ Pascal-Vandermonde matrices, and $\boldsymbol{D}$ is an $r \times r$ block-diagonal matrix

$$\boldsymbol{D} = \mathrm{diag}\left(\boldsymbol{H}_0 \ldots \boldsymbol{H}_{(K-1)}\right)$$

whose k^{th} block $\boldsymbol{H}_k$ is an $M_k \times M_k$ upper anti-triangular Hankel matrix (in the particular case $M_k = 1$, $\boldsymbol{H}_k = z_k^{t-l+1}\alpha_k^{(0)}$).

B. Estimation of the Model Parameters

Proposition 1 shows that the matrix $\boldsymbol{S}$ has rank r, and that its range space, called *signal subspace*, is also spanned by the Pascal-Vandermonde matrix $\boldsymbol{V}^n$.

1) Rotational Invariance Property: The ESPRIT method relies on a particular property of Vandermonde matrices known as *rotational invariance* [1], which reflects the invariance of the signal subspace to time shifts. Theorem 2, shown in [2], generalizes this property to Pascal-Vandermonde matrices.

Theorem 2 (Rotational Invariance Property of Pascal-Vandermonde Matrices): Let $\boldsymbol{V}_{\downarrow}^n$ be the matrix extracted from $\boldsymbol{V}^n$ by deleting the last row. Similarly, let $\boldsymbol{V}_{\uparrow}^n$ be the matrix extracted from $\boldsymbol{V}^n$ by deleting the first row. Then $\boldsymbol{V}_{\downarrow}^n$ and $\boldsymbol{V}_{\uparrow}^n$ span the same subspace, and

$$\boldsymbol{V}_{\uparrow}^n = \boldsymbol{V}_{\downarrow}^n \boldsymbol{J} \tag{5}$$

where $\boldsymbol{J}$ is the $r \times r$ block-diagonal matrix

$$\boldsymbol{J} = \mathrm{diag}\left(\boldsymbol{J}_0, \ldots, \boldsymbol{J}_{(K-1)}\right) \tag{6}$$

whose k^{th} block $\boldsymbol{J}_k$ is the $M_k \times M_k$ Jordan block whose diagonal coefficients are equal to z_k.

The interesting fact in theorem 2 is that (5) involves a Jordan matrix[4] $\boldsymbol{J}$, which characterizes the poles z_k and their multiplicity M_k. As shown below, the generalized ESPRIT algorithm consists in computing $\boldsymbol{J}$ as a byproduct of the Jordan canonical decomposition of a so-called *spectral matrix*.

2) The Generalized ESPRIT Method: In practice, the Pascal-Vandermonde matrix $\boldsymbol{V}^n$ is unknown. Nevertheless, it was shown in [2] that in presence of white noise an $n \times r$ orthonormal matrix $\boldsymbol{W}$ spanning the signal subspace can be estimated by computing the left dominant r-dimensional singular subspace of the noisy data matrix, or by using subspace tracking methods [29]–[31]. Since $\boldsymbol{W}$ and $\boldsymbol{V}^n$ span the same subspace, there is an $r \times r$ invertible matrix $\boldsymbol{G}$ such that

$$\boldsymbol{V}^n = \boldsymbol{W}\boldsymbol{G}. \tag{7}$$

Substituting (7) into (5) shows that $\boldsymbol{W}$ satisfies an equation similar to (5): $\boldsymbol{W}_{\uparrow} = \boldsymbol{W}_{\downarrow}\boldsymbol{\Phi}$ where $\boldsymbol{\Phi}$, herein called the *spectral matrix*, is defined by its Jordan canonical decomposition

$$\boldsymbol{\Phi} \triangleq \boldsymbol{G}\boldsymbol{J}\boldsymbol{G}^{-1}. \tag{8}$$

It can be noticed that $\forall t \in \mathbb{Z}$, the spectral matrix $\boldsymbol{\Phi}$, which depends on the observation window $\{t-l+1, \ldots, t+n-1\}$, is similar to the unique Jordan matrix $\boldsymbol{J}$. Finally, the generalized ESPRIT algorithm consists in the following:

- estimating a basis $\boldsymbol{W}$ of the signal subspace, via singular value decomposition for instance,[5]
- computing the spectral matrix[6] $\boldsymbol{\Phi} = \boldsymbol{W}_{\downarrow}^{\dagger}\boldsymbol{W}_{\uparrow}$.
- computing the eigenvalues of $\boldsymbol{\Phi}$ from which the estimated poles and their multiplicities can be extracted.

Note that in a noisy context, the estimated spectral matrix does not have multiple eigenvalues in practice, and the generalized ESPRIT algorithm cannot be applied as it is. This problem will be discussed in Section A.3) of the Appendix.

3) Estimation of the Complex Amplitudes: The complex amplitudes are estimated by means of the LS method. Let $\widehat{\boldsymbol{V}}^N$ be the $N \times r$ Pascal-Vandermonde matrix defined from the estimated poles, and $\widehat{\boldsymbol{s}}$ the N-dimensional vector containing the successive samples of the observed signal. Then the LS-estimate of the vector $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_0, \ldots, \boldsymbol{\alpha}_{K-1}]^T$ (with $\boldsymbol{\alpha}_k = [\alpha_k^{(0)}, \ldots, \alpha_k^{(M_k-1)}]^T$) containing the complex amplitudes is

$$\widehat{\boldsymbol{\alpha}} = \widehat{\boldsymbol{V}}^{N\dagger}\widehat{\boldsymbol{s}}. \tag{9}$$

III. Perturbation Analysis

The objective here is to measure the performance of the estimators presented above in terms of bias and dispersion. Unfortunately, it is not possible to establish analytic formulae in the general case, because of the eigen or singular value decompositions. However, asymptotic results could be obtained by using the perturbation theory in the case of the sinusoidal model [23] and in the case of the ESM model [24], under the hypothesis of a high SNR. We propose to apply the perturbation theory in the more general framework of the PACE model, in order to finally derive the first and second moments of the estimators. First, we analyze the perturbation induced onto the frequencies and damping factors, from which we derive the perturbation induced onto the amplitudes and phases. The detailed mathematical developments can be found in the Appendix.

Suppose that the PACE signal $s(t)$ is corrupted by a perturbation $\varepsilon\Delta s(t)$ (where $\varepsilon \ll 1$), so that the observed signal is $s(t) + \varepsilon\Delta s(t)$. In other terms, the $n \times l$ Hankel data matrix $\boldsymbol{S}$ is corrupted by an additive perturbation $\varepsilon\Delta\boldsymbol{S}$, where $\Delta\boldsymbol{S}$ is the $n \times l$ Hankel matrix containing the samples of $\Delta s(t)$, so that the observed matrix is

$$\boldsymbol{S}(\varepsilon) = \boldsymbol{S} + \varepsilon\Delta\boldsymbol{S}. \tag{10}$$

Then suppose that the generalized ESPRIT algorithm is applied to the perturbed matrix $\boldsymbol{S}(\varepsilon)$ instead of the exact matrix $\boldsymbol{S}$. In Section A in the Appendix, it is shown that the perturbed subspace weighting matrix $\boldsymbol{W}(\varepsilon)$ (Section A.1), spectral matrix $\boldsymbol{\Phi}(\varepsilon)$ (Section A.2), poles $z_k(\varepsilon)$, frequencies $f_k(\varepsilon)$ and damping factors $\delta_k(\varepsilon)$ (Section A.3) are $\mathcal{C}^\infty$ functions of ε in the neighborhood of $\varepsilon = 0$, leading to the first-order expansions in proposition 3.

[4]See [28, pp. 121–142] for a definition of Jordan canonical decomposition.

[5]In linear systems realization theory, Ho's algorithm is a well-known method for identifying a state space representation [27]. The use of the singular value decomposition in this context was early proposed in [32].

[6]In the whole paper, the symbol † denotes the Moore-Penrose pseudo-inverse.

Proposition 3 (Perturbation of the Frequencies and Damping Factors): Let[7] $\delta_k(\varepsilon) = \Re(\ln(z_k(\varepsilon)))$ and $f_k(\varepsilon) = (1/2\pi)\Im(\ln(z_k(\varepsilon)))$. Then the functions $\varepsilon \mapsto \delta_k(\varepsilon)$ and $\varepsilon \mapsto f_k(\varepsilon)$ are $\mathcal{C}^\infty$ and admit the first-order expansions:

$$\begin{cases} \delta_k(\varepsilon) = \delta_k + \varepsilon\Delta\delta_k + O(\varepsilon^2) \\ f_k(\varepsilon) = f_k + \varepsilon\Delta f_k + O(\varepsilon^2). \end{cases} \tag{11}$$

The first-order perturbations $\Delta\delta_k$ and Δf_k are of the form

$$\begin{cases} \Delta\delta_k = \frac{1}{M_k}\Re\left(\frac{\boldsymbol{u}_k{}^H\Delta\boldsymbol{s}}{z_k\alpha_k^{(M_k-1)}}\right) \\ \Delta f_k = \frac{1}{2\pi M_k}\Im\left(\frac{\boldsymbol{u}_k{}^H\Delta\boldsymbol{s}}{z_k\alpha_k^{(M_k-1)}}\right), \end{cases} \tag{12}$$

where the vectors $\Delta\boldsymbol{s} = [\Delta s(t-l+1), \ldots, \Delta s(t+n-1)]^T$ and $\boldsymbol{u}_k$ (whose coefficients, defined in (45) in the Appendix, depend on the model parameters) have dimension $N = n + l - 1$.

Knowing the influence of a perturbation of the data onto the estimated poles, it is then possible to analyze the perturbation induced onto the amplitudes and phases, obtained from these poles by the least squares method. More precisely, it is shown in Section B.2 in the Appendix that the complex amplitudes $\alpha_k^{(m)}(\varepsilon)$, real amplitudes $a_k^{(m)}(\varepsilon)$ and phases $\phi_k^{(m)}(\varepsilon)$ are $\mathcal{C}^\infty$ functions of ε in the neighborhood of $\varepsilon = 0$, leading to the first-order expansions in proposition 4.

Proposition 4 (Perturbation of the Amplitudes and Phases): Let $a_k^{(m)}(\varepsilon) = |\alpha_k^{(m)}(\varepsilon)|$ and $\phi_k^{(m)}(\varepsilon) = (1/2\pi)\Im(\ln(\alpha_k^{(m)}(\varepsilon)))$. If $a_k^{(m)} \neq 0$, then the functions $\varepsilon \mapsto a_k^{(m)}(\varepsilon)$ and $\varepsilon \mapsto \phi_k^{(m)}(\varepsilon)$ are $\mathcal{C}^\infty$ in the neighborhood of $\varepsilon = 0$, and admit the first-order expansion

$$\begin{cases} a_k^{(m)}(\varepsilon) = a_k^{(m)} + \varepsilon\Delta a_k^{(m)} + O(\varepsilon^2) \\ \phi_k^{(m)}(\varepsilon) = \phi_k^{(m)} + \varepsilon\Delta\phi_k^{(m)} + O(\varepsilon^2). \end{cases} \tag{13}$$

The first-order perturbations $\Delta a_k^{(m)}$ and $\Delta\phi_k^{(m)}$ are of the form

$$\begin{cases} \Delta a_k^{(m)} = a_k^{(m)}\Re\left(\frac{\boldsymbol{b}_k^{(m)H}\Delta\boldsymbol{s}}{\alpha_k^{(m)}}\right) \\ \Delta\phi_k^{(m)} = \Im\left(\frac{\boldsymbol{b}_k^{(m)H}\Delta\boldsymbol{s}}{\alpha_k^{(m)}}\right) \end{cases} \tag{14}$$

where the vectors $\boldsymbol{b}_k^{(m)}$ (whose coefficients, defined in Section B in the Appendix, depend on the model parameters) have dimension N.

IV. Performance of the Estimators

Here we aim at exploiting the results of the perturbation analysis in Section III to derive the first and second moments of the generalized ESPRIT estimators [2] in the case of a high SNR. The most remarkable property shown below is that the asymptotic efficiency of the estimators of all parameters associated to single undamped poles is minimum if and only if the number of rows n and the number of columns l of the data matrix satisfy either $n = 2l$ or $l = 2n$. In practice, this result allows to properly dimension the data matrix when the length of the observation window $N = n + l - 1$ is fixed.

From now on, we suppose that $\Delta s(t)$ is a circular complex stationary process of variance 1. Then the stationary process $\varepsilon\Delta s(t)$ can be viewed as a complex noise of variance $\sigma^2 = \varepsilon^2 \ll 1$. Using vector notations, we can write $\boldsymbol{s}(\varepsilon) = \boldsymbol{s} + \varepsilon\Delta\boldsymbol{s}$, where the vectors $\boldsymbol{s}(\varepsilon)$, $\boldsymbol{s}$ and $\Delta\boldsymbol{s}$ contain the N successive samples of the corresponding signals in the observation window $\{t-l+1 \ldots t+n-1\}$. The covariance matrix $\boldsymbol{\Gamma} \triangleq \mathbb{E}[\Delta\boldsymbol{s}\Delta\boldsymbol{s}^H]$ of the circular complex random vector $\Delta\boldsymbol{s}$ is an $N \times N$ Toeplitz matrix, whose diagonal coefficients are equal to 1. In the particular case of white noise, $\boldsymbol{\Gamma}$ is equal to the identity.

Section IV-A presents the first-order performance analysis, which is then simplified in Section IV-B in an asymptotic context. The proofs of the results presented in this Section can be found in the Appendix.

A. First-Order Performance

The following proposition gives the variances of the estimated poles $\widehat{z}_k$, damping (or amplifying) factors $\widehat{\delta}_k$ and frequencies $\widehat{f}_k$, obtained by applying the generalized ESPRIT algorithm [2] to the perturbed signal $\widehat{\boldsymbol{s}} = \boldsymbol{s}(\varepsilon)$.

Proposition 5: The estimator $\widehat{z}_k = z_k(\varepsilon)$ of the pole z_k is unbiased at the first order, and its variance is of the form

$$\mathrm{var}(\widehat{z}_k) \sim \frac{\sigma^2}{\left(M_k a_k^{(M_k-1)}\right)^2}\boldsymbol{u}_k{}^H\boldsymbol{\Gamma}\boldsymbol{u}_k. \tag{15}$$

where the vector $\boldsymbol{u}_k$ is defined in (45). In the same way, the estimators $\widehat{\delta}_k = \delta_k(\varepsilon)$ and $\widehat{f}_k = f_k(\varepsilon)$ of the damping factor δ_k and of the frequency f_k are unbiased at the first order. Moreover, their respective variances are of the form

$$\mathrm{var}(\widehat{\delta}_k) \sim \frac{\sigma^2 e^{-2\delta_k}}{2\left(M_k a_k^{(M_k-1)}\right)^2}\boldsymbol{u}_k{}^H\boldsymbol{\Gamma}\boldsymbol{u}_k \tag{16}$$

$$\mathrm{var}(\widehat{f}_k) \sim \frac{1}{4\pi^2}\mathrm{var}(\widehat{\delta}_k). \tag{17}$$

The variances of $\widehat{\delta}_k$ and $\widehat{f}_k$ are derived from their first-order expansions, presented in Proposition 3 in Section III. A remarkable similitude can be noticed between their expressions and the CRB, whose formulas were derived in (10) and (11) in [12].[8] However, (16) and (17) are not self-explanatory, because they involve many auxiliary variables, via the vectors $\boldsymbol{u}_k$. It will be shown in Section IV-B that under some additional hypotheses, they can be simplified.

These variances satisfy the following properties[9]:

- if the noise is white ($\boldsymbol{\Gamma} = \boldsymbol{I}_N$), the variances of $\widehat{\delta}_k$ and $\widehat{f}_k$ depend on the frequencies only by their differences[10];
- if z_k is a single pole, the variances of $\widehat{\delta}_k$ and $\widehat{f}_k$ do not depend on any phase, and they are inversely proportional to $a_k^{(0)^2}$, but they do not depend on any other amplitude.[11]

Proposition 6 below gives the variances of the estimated complex amplitudes $\widehat{\boldsymbol{\alpha}}$, real amplitudes $\widehat{a}_k^{(m)}$ and phases $\widehat{\phi}_k^{(m)}$, obtained by means of the least squares (LS) method [2].

[7] It is supposed here that all frequencies lie between $-1/2$ and 1/2.

[8] The expressions of the CRB of the PACE model were also presented in Proposition III.2.1 in [38, p. 37].

[9] These properties can be proved by a thorough analysis of the particular structure of the vector $\boldsymbol{u}_k$, defined in (45).

[10] This property is also satisfied by the CRB [12].

[11] In the case of the CRB, this property also applies to multiple poles [12].

Proposition 6: The estimator $\widehat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}(\varepsilon)$ of the vector of complex amplitudes $\boldsymbol{\alpha}$ is unbiased at first order, and its covariance matrix is of the form

$$\mathrm{cov}(\widehat{\boldsymbol{\alpha}}) \sim \sigma^2 \boldsymbol{B}^H \boldsymbol{\Gamma} \boldsymbol{B} \tag{18}$$

where the $N \times r$ matrix $\boldsymbol{B}$ is defined in (53). In the same way, the estimators $\widehat{a}_k^{(m)} = a_k^{(m)}(\varepsilon)$ and $\widehat{\phi}_k^{(m)} = \phi_k^{(m)}(\varepsilon)$ of the real-valued amplitude $a_k^{(m)}$ and the phase $\phi_k^{(m)}$ are unbiased at first order. Moreover, their respective variances are of the form

$$\mathrm{var}\left(\widehat{a}_k^{(m)}\right) \sim \frac{\sigma^2}{2} \boldsymbol{b}_k^{(m)H} \boldsymbol{\Gamma} \boldsymbol{b}_k^{(m)} \tag{19}$$

$$\mathrm{var}\left(\widehat{\phi}_k^{(m)}\right) \sim \frac{1}{{a_k^{(m)}}^2} \mathrm{var}\left(\widehat{a}_k^{(m)}\right), \tag{20}$$

where the vectors $\boldsymbol{b}_k^{(m)}$, introduced in proposition 4, are the columns of the matrix $\boldsymbol{B}$.

As for proposition 5, the variances of $\widehat{\alpha}_k$, $\widehat{a}_k$ and $\widehat{\phi}_k$ are obtained from their first-order expansions, presented in Proposition 4 in Section III.

B. Asymptotic Performance

Let us suppose that the noise is white ($\boldsymbol{\Gamma} = \boldsymbol{I}_N$) and that all poles are on the unit circle ($\forall k \in \{0 \dots K-1\}$, $\delta_k \to 0$). We present below some asymptotic expansions of the estimator variances with respect to n, l and N.

Corollary 7: If z_k is a single pole, the variances of the estimators $\widehat{\delta}_k$ and $\widehat{f}_k$ admit the expansions

$$\mathrm{var}(\widehat{\delta}_k) \sim \frac{\sigma^2}{\max(n,l)^2 \min(n,l) {a_k^{(0)}}^2} + O\left(\frac{1}{N^4}\right) \tag{21}$$

$$\mathrm{var}(\widehat{f}_k) \sim \frac{\sigma^2}{4\pi^2 \max(n,l)^2 \min(n,l) {a_k^{(0)}}^2} + O\left(\frac{1}{N^4}\right). \tag{22}$$

Both of them are minimal for $n = 2l = (2/3)(N+1)$ or for $l = 2n = (2/3)(N+1)$ (if N equals 2 modulo 3), and these minima admit the asymptotic expansions

$$\mathrm{var}(\widehat{\delta}_k) \sim \frac{27}{4} \frac{\sigma^2}{N^3 {a_k^{(0)}}^2}$$

$$\mathrm{var}(\widehat{f}_k) \sim \frac{27}{4} \frac{\sigma^2}{4\pi^2 N^3 {a_k^{(0)}}^2}.$$

The proof of corollary 7 relies on the first-order expansion of the vector $\boldsymbol{u}_k$ involved in (16) and (17), which admits a simple closed form. Attention must be paid to the fact that expressions (21) and (22) are only valid for a single pole. If z_k is a multiple pole, these variances cannot be formulated in such a simple way and are function of the complex amplitudes associated to z_k for all indices $m \in \{0 \dots M_k - 1\}$ (the optimal values of n and l are also function of these amplitudes in this case). These variances can be compared to the asymptotic CRB derived in (14) and (16) in [12].[12]

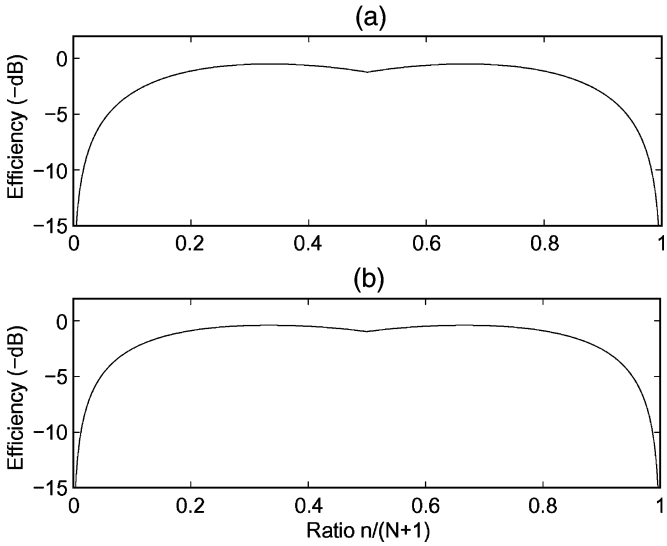

Fig. 1. Efficiency of the estimators. (a) Efficiency of the estimation of the frequencies and damping factors; (b) efficiency of the estimation of the real-valued amplitudes and phases.

> Under the above hypotheses, the asymptotic efficiency of the estimators of all damping factors and all frequencies associated to the single poles is the same one, independent from the model parameters, and equal to $9/8 = 1.125$ if $n = 2l$ or $l = 2n$.

In this way, the results obtained in [19] about the Matrix Pencil method (in the particular case of a single complex sinusoid) are recovered. Fig. 1(a) represents the ratio between the CRB and the variance of the estimators in a logarithmic scale, as a function of the ratio $n/(N+1)$. Thus, it can be verified that the maximum is reached at $n = (N+1)/3$ and $n = 2(N+1)/3$ as expected. Besides, the performance collapses when n becomes too high or too small.

The following corollary is the analogue of corollary 7 for the real-valued amplitudes and phases.

Corollary 8: If z_k is a single pole, the variances of the estimators $\widehat{a}_k^{(0)}$ and $\widehat{\phi}_k^{(0)}$ admit the expansions

$$\mathrm{var}\left(\widehat{a}_k^{(0)}\right) \sim \frac{\sigma^2}{2}\left(\frac{1}{N} + \frac{N^2}{2\max(n,l)^2 \min(n,l)}\right) + O\left(\frac{1}{N^2}\right) \tag{23}$$

$$\mathrm{var}\left(\widehat{\phi}_k^{(0)}\right) \sim \frac{\sigma^2}{2{a_k^{(0)}}^2}\left(\frac{1}{N} + \frac{N^2}{2\max(n,l)^2 \min(n,l)}\right) + O\left(\frac{1}{N^2}\right). \tag{24}$$

[12]The expressions of the asymptotic CRB of the PACE model were also presented in Proposition III.2.2 in [38, p. 38].

Both of them are minimal for $n = 2l = (2/3)(N+1)$ or for $l = 2n = (2/3)(N+1)$ (if N equals 2 modulo 3), and these minima admit the asymptotic expansions

$$\mathrm{var}\left(\widehat{a}_k^{(0)}\right) \sim \frac{35}{16}\frac{\sigma^2}{N}$$
$$\mathrm{var}\left(\widehat{\phi}_k^{(0)}\right) \sim \frac{35}{16}\frac{\sigma^2}{N{a_k^{(0)}}^2}.$$

The proof of corollary 8 relies on the first-order expansions of the vectors $\boldsymbol{b}_k^{(m)}$ involved in (19) and (20), which admit a simple closed form. Again, attention must be paid to the fact that expressions (23) and (24) are only valid for a single pole. These variances can be compared to the asymptotic CRB derived in (17) and (18) in [12].[13]

> Under the above hypotheses, the asymptotic efficiency of the estimators of all the real-valued amplitudes and phases associated to single poles is the same one, independent from the model parameters, and equal to $35/32 = 1.09375$ if $n = 2l$ or $l = 2n$.

This efficiency is even better than that of the estimators $\widehat{\delta}_k$ and $\widehat{f}_k$. It can also be noticed that the optimum is obtained for the same values of n and l as in the previous case. Fig. 1(b) represents the ratio between the CRB and the variance of the estimators in a logarithmic scale, as a function of the ratio $n/(N+1)$. Again, the maximum is reached at $n = (N+1)/3$ and $n = 2(N+1)/3$, and the performance collapses when n becomes too high or too small. The similitude between the curves represented in Fig. 1(a) and (b) is noticeable. This could be explained by the fact that the estimation of the amplitudes and the phases directly relies on the estimation of the frequencies and the damping factors.

V. Simulation Results

A. Dependence of the Variances With Respect to the PACE Parameters

This section illustrates the variations of the estimators variances with respect to the parameters of the PACE model. Note that propositions 5, 6 and corollaries 7, 8 show a rather simple dependency on the amplitudes and the variance σ^2. Therefore, we focus here on the dependency on the frequency gap between two components (Section V-A-1), the damping factor (Section V-A-2), the spectral flatness of the noise (Section V-A-3), and the order of a pole (Section V-A-4). For these simulations, the same synthetic signals as those introduced in [12] are used. In the figures, the solid lines represent the theoretical variance of the frequency estimators or that of the damping factor estimators, which are equivalent according to (17). In the same way, the dashed lines represent the theoretical *relative* variance of the amplitude estimators, which is equal to that of the phase estimators, according to (20).

1) Variation of the Variances With Respect to Frequency Gaps: We consider a signal of length $N = 200$, composed of two undamped components ($K = 2$) of same order $M_0 = M_1 = 1$, in white noise ($\boldsymbol{\Gamma} = \boldsymbol{I}_N$ and $\sigma^2 = 1$). These components have zero phases, and same amplitudes, such that $\mathrm{SNR}_0^{(0)} = \mathrm{SNR}_1^{(0)} = 50$ dB. Fig. 2(a) and (b) shows the variations of the variances of the estimators obtained with $n = (2/3)(N+1)$, with respect to the frequency gap $\Delta f = |f_1 - f_0| \in]0, 0.5]$ (f_0 was set to 0). The diamonds represent the empirical variance, obtained by averaging 100 runs of the ESPRIT algorithm. They match the theoretical variance, which confirms the validity of our perturbation analysis for this SNR. The variation rate of the variances is similar to that of the CRB [12]: it is broken at $\Delta f = 1/N = 510^{-3}$, which corresponds to the resolution limit of Fourier analysis. At this limit point, the relative variance of the amplitude estimate is still lower than -60 dB, which shows the good resolution of the ESPRIT algorithm. The efficiencies of both estimators are represented in Fig. 2(c). It can be noticed that they remain close to 1, even when the frequency gap tends to zero.

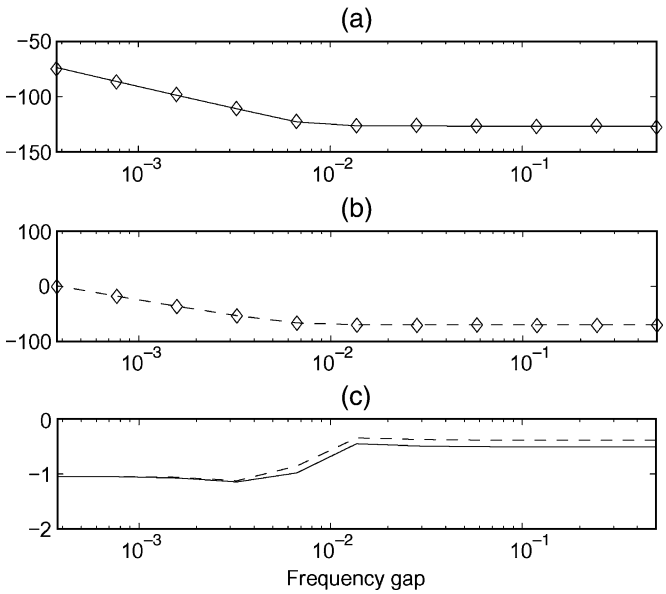

Fig. 2. Variation of the variances with respect to the frequency gap. (a) Variance for the frequencies (dB). (b) Relative variance for the amplitudes (dB). (c) Efficiency (-dB).

2) Variation of the Variances With Respect to the Damping Factor: We consider a signal of length $N = 100$, composed of one component ($K = 1$) of order $M_0 = 1$, in white noise ($\boldsymbol{\Gamma} = \boldsymbol{I}_N$ and $\sigma^2 = 1$). This component has zero frequency and phase, and an amplitude such that $\mathrm{SNR}_0^{(0)} = 50$ dB. Fig. 3(a) and (b) shows the variations of the variances of the estimators obtained with $n = (2/3)(N+1)$, with respect to the damping factor δ_0. The diamonds represent the empirical variance, obtained by averaging 500 runs of the ESPRIT algorithm. They no longer exactly match the theoretical variance for negative values of delta, which shows the validity limit of our perturbation analysis with respect to the high SNR hypothesis. Again, these variations are very similar to those of the CRB illustrated in [12]. The efficiencies of both estimators are represented in Fig. 3(c). They remain close to 1 whatever the value of the damping factor is.

3) Variation of the Variances With Respect to the Spectral Flatness of the Noise: We consider a signal of length $N =$

[13]The expressions of the asymptotic CRB of the PACE model were also presented in Proposition III.2.2. in [38, p. 38].

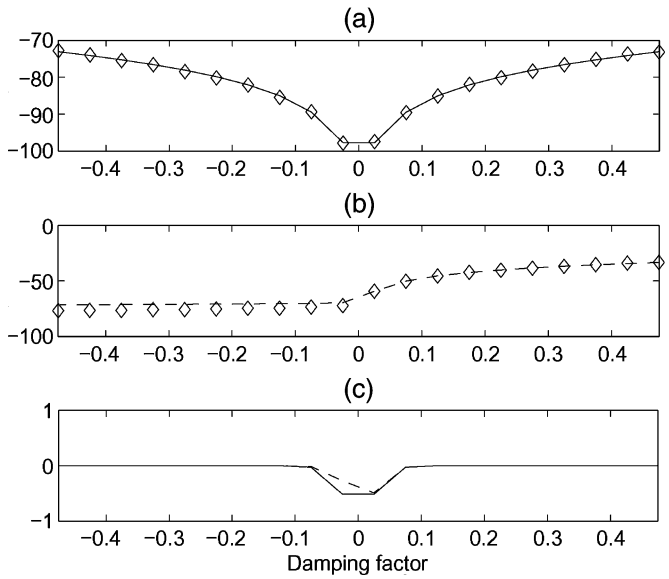

Fig. 3. Variation of the variances with respect to the damping factor. (a) Variance for the damping factor (dB). (b) Relative variance for the amplitude (dB). (c) Efficiency (-dB).

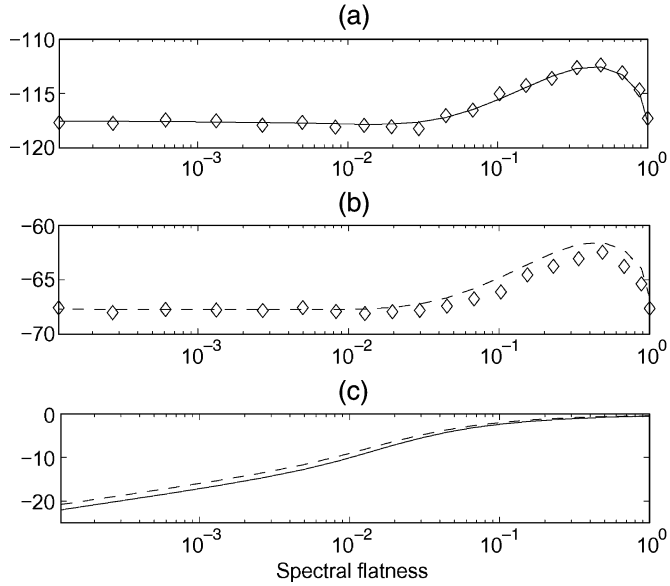

Fig. 4. Variation of the variances with respect to the spectral flatness of the noise. (a) Variance for the frequency (dB). (b) Relative variance for the amplitude (dB). (c) Efficiency (-dB).

100, composed of one undamped component $(K = 1)$ of order $M_0 = 1$, in colored noise. This component has a zero phase, a normalized frequency equal to 0.05, and an amplitude such that $\mathrm{SNR}_0^{(0)} = 50$ dB. The noise is obtained by filtering a white noise by the filter of transfer function $H_a(z) = 1/(1 - az^{-1})$ (where $0 \leq a < 1$), such that $\mathbf{\Gamma} = \mathrm{Toeplitz}(1, a, a^2 \ldots a^{N-1})$. The Spectral Flatness (SF) measure of the noise is defined as

$$\mathrm{SF}(a) = \frac{\exp\left(\int_0^1 \ln\left(\left|H_a(e^{i2\pi f})\right|^2\right) \mathrm{d}f\right)}{\int_0^1 \left|H_a(e^{i2\pi f})\right|^2 \mathrm{d}f}.$$

By tuning the parameter a, it is possible to make the spectral flatness map the range]0,1] (the case $\mathrm{SF} = 1$ corresponds to white noise). Fig. 4 illustrates the variations of the variances of the estimators obtained with $n = (2/3)(N + 1)$, with respect to the spectral flatness of the noise. As expected, Fig. 4(c) shows that the efficiency degrades when the spectral flatness decreases (note that the ESPRIT algorithm explicitly relies on the white noise assumption). In other respects, Fig. 4(a) and (b) shows that the variances admit a maximum when $\mathrm{SF} \simeq 0.5$. At this point, the theoretical and the empirical variances (obtained by averaging 1000 runs) no longer match in Fig. 4(b), which shows the validity limit of our perturbation analysis for this SNR. In the range $\mathrm{SF} \in [0.5, 1]$, the variances decrease when the spectral flatness increases, as expected. In the range $\mathrm{SF} \in]0.02, 0.5]$, we observe the inverse phenomenon. Indeed, as mentioned in [12], the power spectral density of the noise becomes a sharp peak when SF becomes low, and converges to a spectral line when $\mathrm{SF} \to 0$. Therefore, the problem of estimating the single undamped component in colored noise becomes close to the problem of estimating two undamped components *without noise*. However, contrary to what is observed in [12] in the case of the CRB, we note that the variances are stationary in the interval $\mathrm{SF} \in]0, 0.02]$. This is because the ESPRIT algorithm is applied with an "erroneous" model order ($K = 1$ instead of 2).[14]

4) Variation of the Variances With Respect to the Pole Order: We consider a signal of length $N = 20$, composed of one undamped component $(K = 1)$ of order $M_0 \in \{1 \ldots 4\}$, in white noise ($\mathbf{\Gamma} = \boldsymbol{I}_N$ and $\sigma^2 = 1$). This component has zero phases, and amplitudes such that $\mathrm{SNR}_0^{(M_0-1)} = 50$ dB, and $\forall m < M_0 - 1, \mathrm{SNR}_0^{(m)} = 0$. The corresponding pole is $z_0 = 1$. Fig. 5(a) and (b) shows the variations of the variances of the estimators obtained with $n = (2/3)(N + 1)$, with respect to the pole order M_0. They confirm the results obtained for the CRB [12]: estimating multiple poles is all the more difficult as their order is high. Actually this estimation is no longer possible in this case if $M_0 > 4$. Besides, the empirical variance (obtained by averaging 100 runs) and the theoretical variance no longer exactly match for $M_0 \geq 4$, which shows the validity limit of our perturbation analysis for this SNR (we observed that they match again if the SNR becomes greater than 80 dB). In other respects, Fig. 5(c) shows that the efficiency rapidly degrades when M_0 increases. We may infer that the arithmetic mean of the scattered eigenvalues is not a sufficiently reliable estimator for a pole of high multiplicity. Some ideas to improve this estimator are suggested in [34].

B. Application to ARMA Filter Synthesis

As shown in Section V-A-4, estimating multiple poles is a difficult task. In order to illustrate the practical capabilities of the generalized ESPRIT algorithm in presence of multiple poles, we propose below an application to ARMA filter synthesis, in the context of system conversion from continuous time to discrete time.

[14]The impact of an erroneous modeling order on the estimated parameters was studied in [33].

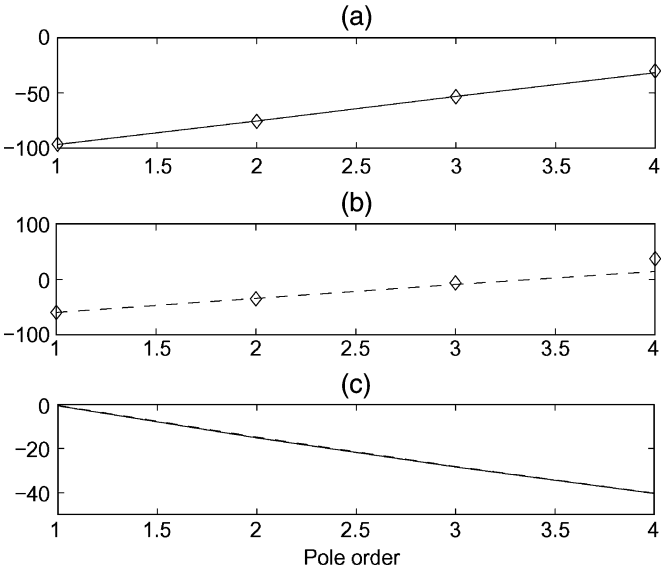

Fig. 5. Variation of the variances with respect to the pole order. (a) Variance for the frequency (dB). (b) Relative variance for the amplitude (dB). (c) Efficiency (-dB).

1) Principle: We consider continuous time systems defined by an ordinary differential equation (ODE) with real-valued coefficients (below τ denotes the continuous time, in seconds)

$$a_0\widetilde{y}(\tau)+a_1\widetilde{y}'(\tau)+\ldots+a_p\widetilde{y}^{(p)}(\tau)=b_0\widetilde{x}(\tau)+\ldots+b_q\widetilde{x}^{(q)}(\tau)$$

whose transfer function can be written in the form

$$\widetilde{H}(\rho)=\frac{\sum_{k=0}^{q} b_k\rho^k}{\sum_{k=0}^{p} a_k\rho^k}$$

where $\rho = i2\pi\nu$ and ν denotes the frequency, in Hertz. Moreover, we suppose that $\widetilde{x}(\tau)$ and $\widetilde{y}(\tau)$ satisfy the conditions of the Nyquist-Shannon sampling theorem: $\exists T>0$ such that the Fourier transforms $\widetilde{X}(\nu)$ and $\widetilde{Y}(\nu)$ are zero outside the range $[-(1/2T),1/2T]$, and we consider the discrete signals $x(t)=\widetilde{x}(tT)$ and $y(t)=\widetilde{y}((t-t_0)T)$ for all $t\in\mathbb{Z}$ (where $t_0\in\mathbb{R}$), obtained by sampling $x(\tau)$ and $y(\tau)$ at the frequency $1/T$. Then it is well known that $y(t)$ can be obtained from $x(t)$ by applying the discrete filter of frequency response

$$H(e^{i2\pi f})=\widetilde{H}\left(i2\pi\frac{f}{T}\right)e^{-i2\pi t_0 f}=\frac{\sum_{k=0}^{q} b_k\left(i2\pi\frac{f}{T}\right)^k}{\sum_{k=0}^{p} a_k\left(i2\pi\frac{f}{T}\right)^k}e^{-i2\pi t_0 f}$$

where $f\in[-(1/2),1/2]$ is the normalized frequency. It can be noticed that the time delay t_0 can be chosen such that $H(e^{i2\pi(1/2)})=H(e^{-i2\pi(1/2)})$. In this case, if the denominator is never zero, the 1-periodic function $f\mapsto H(e^{i2\pi f})$ is continuous and piecewise continuously differentiable, which proves that $h(t)=O(1/t^2)$, thus the discrete filter h is stable. The impulse response $h(t)$ can then be obtained by numerically computing the inverse discrete time Fourier transform of $H(e^{i2\pi f})$. However, this impulse response $h(t)$ is generally infinite, and we aim at approximating it by an AutoRegressive Moving Average (ARMA) filter $g(t)$.

It is well known that the general rational transfer function of a stable ARMA filter can be decomposed in the form

$$G(z)=\sum_{m=0}^{M_0-1}\alpha_0^{(m)}z^{-m}+\sum_{0<|z_k|<1}\sum_{m=0}^{M_k-1}\frac{\alpha_k^{(m)}}{(1-z_kz^{-1})^{(1+m)}}+\sum_{1<|z_k|<+\infty}\sum_{m=0}^{M_k-1}\frac{\beta_k^{(m)}(-z/z_k)^{(1+m)}}{(1-z/z_k)^{(1+m)}} \quad (25)$$

where for all $k\in\{0\ldots K-1\}$ z_k is a pole of multiplicity M_k (here we assume $z_0=0$). This transfer function corresponds to the impulse response

$$g(t)=\sum_{0\le|z_k|<1}\sum_{m=0}^{M_k-1}\alpha_k^{(m)}F_m(t)z_k^{t-m}$$

for all $t\ge 0$, and

$$g(t)=\sum_{1<|z_k|<+\infty}\sum_{m=0}^{M_k-1}(-1)^{(1+m)}\beta_k^{(m)}F_m\left(|t|-1\right)z_k^{-|t|}$$

for all $t<0$.

Thus both the causal and anticausal parts of $g(t)$ satisfy a PACE model.[15] Following this remark, we can find an ARMA filter g which approximates the discrete filter h by applying the generalized ESPRIT algorithm to the impulse response $h(t)$ on two appropriately chosen finite intervals.

2) Example: The ARMA filter synthesis method could be successfully used for designing differentiator or integrator filters. Here it is applied to the continuous time system

$$\widetilde{y}(\tau)+3\widetilde{y}'(\tau)+3\widetilde{y}''(\tau)+\widetilde{y}'''(\tau)=\widetilde{x}(\tau)-\widetilde{x}'(\tau) \quad (26)$$

with parameters $T=1$ and[16] $t_0\simeq 2.3924$. The impulse and the frequency response of the corresponding discrete filter h are represented in Fig. 6.

Equation (26) shows that the continuous time filter contains a triple pole at $\rho=-1$. When synthesizing the corresponding ARMA filter, we thus expect to find a triple pole at $z=e^{\rho/T}=1/e$. Fig. 7 represents the estimated poles of the causal part of g in the complex plane, obtained by applying the ESPRIT algorithm with $N=128$, $n=(N+1)/3=43$, and[17] $r=15$. As expected, we observe a triple pole in the neighborhood of $1/e\simeq 0.3679$ (which is scattered into three single eigenvalues forming the vertices of an equilateral triangle). We also observe a double pole in the neighborhood of 0, which corresponds to the polynomial part of the transfer function in (25). Finally, we obtained an ARMA filter g with 26 poles and 27 zeros. The approximation error for the impulse response was

$$\max_{n\in[-128,128]}|h(n)-g(n)|<5\times 10^{-8}$$

[15]Note that the causal part generally contains a multiple pole $z_0=0$, whereas the anticausal part never contains a pole at $z=\infty$.

[16]The fractional part of the delay t_0 was chosen in order to make the frequency response $H(e^{i2\pi f})$ continuous, and the integer part of t_0 was chosen as the smallest integer such that the anticausal part of the estimated filter g has no pole at $z=\infty$.

[17]The order r was selected by means of Information Theoretic Criteria [35].

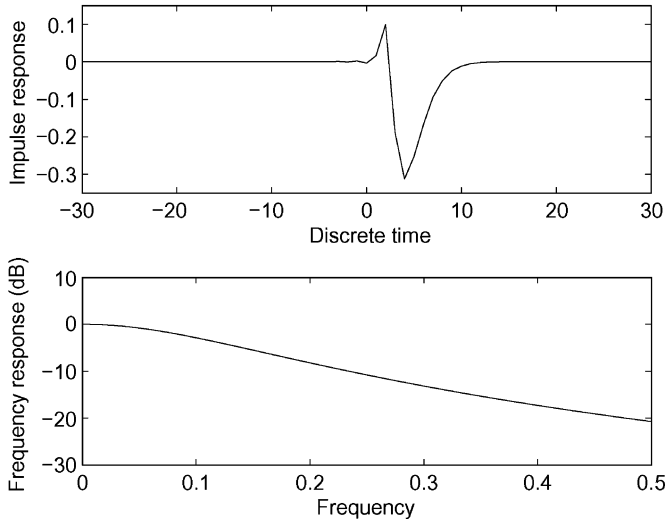

Fig. 6. Impulse and frequency response of the discrete filter.

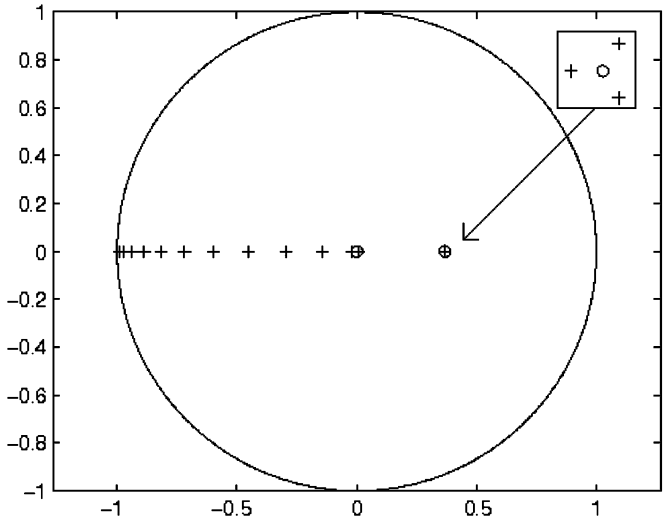

Fig. 7. Poles of the causal part of the ARMA filter.

and the approximation error for the frequency response was

$$\max_{f\in[-(1/2),1/2]} |20\log_{10} H(e^{i2\pi f}) - 20\log_{10} G(e^{i2\pi f})| < 10^{-5}.$$

VI. Conclusion

In this paper, the performance of the generalized ESPRIT algorithm for estimating the parameters of the PACE model, also called quasi-polynomial model, was investigated in the context of high SNRs. This paper was based on the analysis of the first-order perturbations induced by an additive noise. In particular, it was shown that the perturbation of the estimated poles is not sensitive to the particular choice of the orthonormal subspace weighting matrix. In other respects, the presence of noise scatters multiple poles into several single eigenvalues, forming the vertices of a regular polygon. However, it was proved that the estimation of multiple poles can be improved by calculating the arithmetic mean of the scattered eigenvalues.

Then it was shown that the estimators of all the parameters obtained in this way are unbiased, and their variances were calculated and compared to the CRB. By supposing that the noise is white, that all poles are on the unit circle, and that the SNR, n and $l \to +\infty$, it was shown that the efficiency of single poles estimators is close to 1. More precisely, the asymptotic efficiency of the estimators of all damping factors and frequencies is the same one, independent from the model parameters, and equal to $9/8 = 1.125$ if $n = 2l$ or $l = 2n$. In other respects, the asymptotic efficiency of the estimators of all the real-valued amplitudes and phases is the same one, independent from the model parameters, and equal to $35/32 = 1.09375$ if $n = 2l$ or $l = 2n$.

However, our simulation results showed that the variances of the estimators associated to a multiple pole present an exponential increase with the order of the pole. Thus the practical estimation of the PACE model parameters is only possible if the exponentials are modulated by polynomials of low order. Nevertheless, some recent advances in linear algebra computations, such as the techniques proposed in [34], offer interesting outlooks for improving the estimation of multiple poles.

Appendix

The following developments lead to propositions 3 and 4 presented in Section III. The exhaustive proofs of these results are presented in a supporting document [39].

A. Perturbation of the Frequencies and Damping Factors

We successively derive the following first-order expansions:

- $\boldsymbol{W}(\varepsilon) = \boldsymbol{W} + \varepsilon\Delta\boldsymbol{W} + O(\varepsilon^2)$, where the expression of $\Delta\boldsymbol{W}$ is a linear function of $\Delta\boldsymbol{S}$ (proposition 9);
- $\boldsymbol{\Phi}(\varepsilon) = \boldsymbol{\Phi} + \varepsilon\Delta\boldsymbol{\Phi} + O(\varepsilon^2)$, where the expression of $\Delta\boldsymbol{\Phi}$ is a linear function of $\Delta\boldsymbol{W}$ (proposition 10);
- $z_k(\varepsilon) = z_k + \varepsilon\Delta z_k + O(\varepsilon^2)$, where the expression of Δz_k is a linear function of $\Delta\boldsymbol{\Phi}$ (proposition 12);
- $\begin{cases} \delta_k(\varepsilon) = \delta_k + \varepsilon\Delta\delta_k + O(\varepsilon^2) \\ f_k(\varepsilon) = f_k + \varepsilon\Delta f_k + O(\varepsilon^2) \end{cases}$ where the expressions of $\Delta\delta_k$ and Δf_k are functions of Δz_k (corollary 3).

Finally, successive substitutions lead to the expressions of $\Delta\delta_k$ and Δf_k as functions of the additive perturbation Δs.

1) Perturbation of the Signal Subspace: Here we analyze the influence of a perturbation of the data onto the signal subspace. For all ε, let $\boldsymbol{\Pi}(\varepsilon)$ be the $n \times n$ projector onto the r-dimensional dominant subspace of the $n \times n$ positive semidefinite matrix $\boldsymbol{S}(\varepsilon)\boldsymbol{S}(\varepsilon)^H$. If $\boldsymbol{W}$ is orthonormal, $\boldsymbol{\Pi}(0) = \boldsymbol{W}\boldsymbol{W}^H$. Then the perturbation theory shows that the function $\varepsilon \mapsto \boldsymbol{\Pi}(\varepsilon)$ is $\mathcal{C}^\infty$ in a neighborhood of $\varepsilon = 0$. In the literature, the asymptotic performance analysis of some subspace-based algorithms was performed by investigating the perturbation of this projector at the first order [36]. However, we are interested here in the perturbation of the subspace weighting matrix $\boldsymbol{W}$, which is analyzed in the following proposition.

Proposition 9 (Perturbation of the Signal Subspace): There exists an infinity[18] of $\mathcal{C}^\infty$ functions $\varepsilon \mapsto \boldsymbol{W}(\varepsilon)$ defined in a neighborhood of $\varepsilon = 0$ and with values in the group of complex orthonormal matrices $\mathbb{U}^{n\times r}$, which span the r-dimensional dominant subspace of $\boldsymbol{S}(\varepsilon)$ (i.e. such that $\boldsymbol{W}(0) = \boldsymbol{W}$ and

[18]All these functions are obtained by right-multiplying any of them by a $\mathcal{C}^\infty$ function, with values in the group of complex orthonormal matrices $\mathbb{U}^{r\times r}$, and reaching the value $\boldsymbol{I}_r$ at $\varepsilon = 0$.

$\boldsymbol{W}(\varepsilon)\boldsymbol{W}(\varepsilon)^H = \boldsymbol{\Pi}(\varepsilon)$). Each $\mathcal{C}^\infty$ function $\varepsilon \mapsto \boldsymbol{W}(\varepsilon)$ admits a first-order expansion

$$\boldsymbol{W}(\varepsilon) = \boldsymbol{W} + \varepsilon\Delta\boldsymbol{W} + O(\varepsilon^2). \tag{27}$$

The first-order perturbation $\Delta\boldsymbol{W}$ can be decomposed as

$$\Delta\boldsymbol{W} = \Delta\boldsymbol{W}^\perp - \boldsymbol{W}\boldsymbol{A} \tag{28}$$

where the $r \times r$ matrix $\boldsymbol{A} \triangleq \Delta\boldsymbol{W}^H\boldsymbol{W}$ is antihermitian and the $n \times r$ matrix $\Delta\boldsymbol{W}^\perp$ is orthogonal to $\mathrm{span}(\boldsymbol{W})$

$$\Delta\boldsymbol{W}^\perp = (\boldsymbol{I}_n - \boldsymbol{W}\boldsymbol{W}^H)\Delta\boldsymbol{S}\boldsymbol{S}^\dagger\boldsymbol{W}. \tag{29}$$

The proof of proposition 9 relies on the implicit definition of $\boldsymbol{W}(\varepsilon)$ as the unique minimum point of a cost function $\mathcal{J}$. Equations (27) to (29) are thus derived by zeroing the first derivative of $\mathcal{J}$.

Proof of proposition 9: It can be verified that the function

$$\boldsymbol{W}(\varepsilon) \triangleq \boldsymbol{\Pi}(\varepsilon)\boldsymbol{W}\left(\boldsymbol{W}^H\boldsymbol{\Pi}(\varepsilon)\boldsymbol{W}\right)^{-\frac{1}{2}}$$

satisfies all the properties mentioned in proposition 9. It is also clear that any function of the form $\boldsymbol{W}(\varepsilon)\boldsymbol{\Theta}(\varepsilon)$, where $\varepsilon \mapsto \boldsymbol{\Theta}(\varepsilon)$ is a $\mathcal{C}^\infty$ function, whose values belong to the group of orthonormal matrices $\mathcal{O}_r(\mathbb{C})$, and which reaches the value $\boldsymbol{I}_r$ at $\varepsilon = 0$, also satisfies these properties. Lastly, if $\varepsilon \mapsto \boldsymbol{W}'(\varepsilon)$ is an other function satisfying all these properties, then $\boldsymbol{\Pi}(\varepsilon) = \boldsymbol{W}(\varepsilon)\boldsymbol{W}(\varepsilon)^H = \boldsymbol{W}'(\varepsilon)\boldsymbol{W}'(\varepsilon)^H$. Therefore, $\boldsymbol{W}'(\varepsilon) = \boldsymbol{W}(\varepsilon)\boldsymbol{\Theta}(\varepsilon)$, where $\boldsymbol{\Theta}(\varepsilon) \triangleq \boldsymbol{W}(\varepsilon)^H\boldsymbol{W}'(\varepsilon)$ is a $\mathcal{C}^\infty$ function, whose values belong to the group of orthonormal matrices $\mathcal{O}_r(\mathbb{C})$ since $\boldsymbol{W}(\varepsilon)$ and $\boldsymbol{W}'(\varepsilon)$ are two orthonormal bases of the same subspace, and which reach the value $\boldsymbol{W}^H\boldsymbol{W} = \boldsymbol{I}_r$ at $\varepsilon = 0$. Then note that according to [37], any orthonormal matrix $\boldsymbol{W}(\varepsilon)$ spanning the principal subspace of the matrix $\boldsymbol{S}(\varepsilon)\boldsymbol{S}(\varepsilon)^H$ minimizes the function

$$\begin{array}{rccc} \mathcal{J}: & \mathbb{C}^{n\times r} & \to & \mathbb{R} \\ & \boldsymbol{U} & \mapsto & \frac{1}{2}\left\|\boldsymbol{S}(\varepsilon) - \boldsymbol{U}\boldsymbol{U}^H\boldsymbol{S}(\varepsilon)\right\|_F^2. \end{array}$$

Consequently, the gradient $\boldsymbol{\nabla}\mathcal{J}(\boldsymbol{U})$ is zero at $\boldsymbol{U} = \boldsymbol{W}(\varepsilon)$. However, it can be verified[19] that $\boldsymbol{\nabla}\mathcal{J}(\boldsymbol{U}) =$

$$\left(-2\boldsymbol{S}(\varepsilon)\boldsymbol{S}(\varepsilon)^H + \boldsymbol{S}(\varepsilon)\boldsymbol{S}(\varepsilon)^H\boldsymbol{U}\boldsymbol{U}^H + \boldsymbol{U}\boldsymbol{U}^H\boldsymbol{S}(\varepsilon)\boldsymbol{S}(\varepsilon)^H\right)\boldsymbol{U}.$$

Let $\boldsymbol{W}(\varepsilon) = \boldsymbol{W} + \varepsilon\Delta\boldsymbol{W} + O(\varepsilon^2)$ be the first-order expansion of the function $\varepsilon \mapsto \boldsymbol{W}(\varepsilon)$. Then

$$\boldsymbol{\nabla}\mathcal{J}\left(\boldsymbol{W}(\varepsilon)\right) = \varepsilon\boldsymbol{\nabla}\mathcal{J}^{(1)} + O(\varepsilon^2) \tag{30}$$

where

$$\boldsymbol{\nabla}\mathcal{J}^{(1)} = -(\boldsymbol{I}_n - \boldsymbol{W}\boldsymbol{W}^H)\Delta\boldsymbol{S}\boldsymbol{S}^H\boldsymbol{W} + \Delta\boldsymbol{W}(\boldsymbol{W}^H\boldsymbol{S}\boldsymbol{S}^H\boldsymbol{W}) + \boldsymbol{W}\boldsymbol{N} \tag{31}$$

[19]To compute $\boldsymbol{\nabla}\mathcal{J}(\boldsymbol{U})$, the following derivation rule has to be applied: if $\boldsymbol{M}$ is a constant $n \times r$ matrix, $\boldsymbol{\nabla}\mathrm{trace}(\Re(\boldsymbol{U}^H\boldsymbol{M})) = \boldsymbol{M}$. As a consequence, if $\boldsymbol{C}$ is a constant $n \times n$ hermitian matrix, $\boldsymbol{\nabla}\mathrm{trace}(\boldsymbol{U}^H\boldsymbol{C}\boldsymbol{U}) = 2\boldsymbol{C}\boldsymbol{U}$.

and

$$\boldsymbol{N} \triangleq (\boldsymbol{W}^H\boldsymbol{S}\boldsymbol{S}^H\boldsymbol{W})(\boldsymbol{W}^H\Delta\boldsymbol{W} + \Delta\boldsymbol{W}^H\boldsymbol{W}) + \Delta\boldsymbol{W}^H\boldsymbol{W}(\boldsymbol{W}^H\boldsymbol{S}\boldsymbol{S}^H\boldsymbol{W}). \tag{32}$$

However, the first-order expansion of the orthonormality condition $\boldsymbol{W}(\varepsilon)^H\boldsymbol{W}(\varepsilon) = \boldsymbol{I}_r$ shows that $\boldsymbol{W}^H\Delta\boldsymbol{W} + \Delta\boldsymbol{W}^H\boldsymbol{W} = \boldsymbol{0}_{(r\times r)}$, which means that the matrix $\boldsymbol{A} \triangleq \Delta\boldsymbol{W}^H\boldsymbol{W}$ is antihermitian. Thus $\boldsymbol{N} = \boldsymbol{A}(\boldsymbol{W}^H\boldsymbol{S}\boldsymbol{S}^H\boldsymbol{W})$. Since $\boldsymbol{\nabla}\mathcal{J}(\boldsymbol{W}(\varepsilon)) = \boldsymbol{0}_{(n\times r)}$, (30) to (32) yield

$$\Delta\boldsymbol{W} = (\boldsymbol{I}_n - \boldsymbol{W}\boldsymbol{W}^H)\Delta\boldsymbol{S}\boldsymbol{S}^H\boldsymbol{W}(\boldsymbol{W}^H\boldsymbol{S}\boldsymbol{S}^H\boldsymbol{W})^{-1} - \boldsymbol{W}\boldsymbol{A}.$$

Finally, by noting that $\boldsymbol{S}^H\boldsymbol{W}(\boldsymbol{W}^H\boldsymbol{S}\boldsymbol{S}^H\boldsymbol{W})^{-1} = \boldsymbol{S}^\dagger\boldsymbol{W}$, (27) to (29) can be derived. □

2) Perturbation of the Spectral Matrix: The following proposition complements the result of proposition 9 by showing how the spectral matrix is perturbed.

Proposition 10 (Perturbation of the Spectral Matrix): Suppose that the matrix $\boldsymbol{W}_\downarrow$ is full-rank. Then in the neighborhood of $\varepsilon = 0$, $\boldsymbol{W}_\downarrow(\varepsilon)$ is also full-rank. Moreover, the function

$$\boldsymbol{\Phi}(\varepsilon) \triangleq \boldsymbol{W}(\varepsilon)_\downarrow^\dagger\boldsymbol{W}(\varepsilon)_\uparrow \tag{33}$$

is $\mathcal{C}^\infty$, and admits the first-order expansion

$$\boldsymbol{\Phi}(\varepsilon) = \boldsymbol{\Phi} + \varepsilon\Delta\boldsymbol{\Phi} + O(\varepsilon^2). \tag{34}$$

The first-order perturbation $\Delta\boldsymbol{\Phi}$ can be written in the form

$$\Delta\boldsymbol{\Phi} = \Delta\boldsymbol{\Phi}^\perp + \boldsymbol{A}\boldsymbol{\Phi} - \boldsymbol{\Phi}\boldsymbol{A} \tag{35}$$

where the $r \times r$ matrix $\Delta\boldsymbol{\Phi}^\perp$ is defined as

$$\Delta\boldsymbol{\Phi}^\perp = -\boldsymbol{W}_\downarrow^\dagger\Delta\boldsymbol{W}_\downarrow^\perp\boldsymbol{\Phi} + \boldsymbol{\Phi}\boldsymbol{W}_\uparrow^\dagger\Delta\boldsymbol{W}_\uparrow^\perp. \tag{36}$$

Equations (34) to (36) are obtained by substituting (27) to (29) into the first-order expansion of (33) which defines $\boldsymbol{\Phi}(\varepsilon)$.

In the following step, the estimated poles are defined as the eigenvalues of the perturbed spectral matrix $\boldsymbol{\Phi}(\varepsilon)$. In order to compute their first-order expansion, we first need to introduce the matrix $\boldsymbol{J}(\varepsilon) \triangleq \boldsymbol{G}^{-1}\boldsymbol{\Phi}(\varepsilon)\boldsymbol{G}$, which is similar to $\boldsymbol{\Phi}(\varepsilon)$; thus the estimated poles can also be viewed as the eigenvalues of $\boldsymbol{J}(\varepsilon)$. Note, however, that $\boldsymbol{J}(\varepsilon)$ is generally no longer a Jordan matrix when $\varepsilon > 0$. The following corollary provides the first-order expansion of this matrix. Let us define the vectors $\boldsymbol{v}_\uparrow$ and $\boldsymbol{v}_\downarrow$ of same dimension r as the conjugate transpose of the first and last row of the matrix $\boldsymbol{V}^n$, respectively, and consider the $r \times r$ positive definite matrix $\boldsymbol{Z} = \boldsymbol{V}^{nH}\boldsymbol{V}^n$.

Corollary 11: Let $\boldsymbol{J}(\varepsilon) = \boldsymbol{G}^{-1}\boldsymbol{\Phi}(\varepsilon)\boldsymbol{G}$. The function $\varepsilon \mapsto \boldsymbol{J}(\varepsilon)$ is $\mathcal{C}^\infty$ in a neighborhood of $\varepsilon = 0$, and admits the expansion

$$\boldsymbol{J}(\varepsilon) = \boldsymbol{J} + \varepsilon\Delta\boldsymbol{J} + O(\varepsilon^2) \tag{37}$$

where the first-order perturbation $\Delta\boldsymbol{J} = \boldsymbol{G}^{-1}\Delta\boldsymbol{\Phi}\boldsymbol{G}$ can be written in the following form (where $\boldsymbol{A}' = \boldsymbol{G}^{-1}\boldsymbol{A}\boldsymbol{G}$):

$$\Delta\boldsymbol{J} = \Delta\boldsymbol{J}^\perp + \boldsymbol{A}'\boldsymbol{J} - \boldsymbol{J}\boldsymbol{A}'. \tag{38}$$

Moreover, the matrix $\Delta \boldsymbol{J}^{\perp} = \boldsymbol{G}^{-1}\Delta\boldsymbol{\Phi}^{\perp}\boldsymbol{G}$ has rank 2

$$\Delta \boldsymbol{J}^{\perp} = \boldsymbol{v}'_{\downarrow}\boldsymbol{e}_{\downarrow}{}^{H}\Delta\boldsymbol{S}\boldsymbol{V}^{l\dagger T}\boldsymbol{D}^{-1}\boldsymbol{J} - \boldsymbol{J}\boldsymbol{v}'_{\uparrow}\boldsymbol{e}_{\uparrow}{}^{H}\Delta\boldsymbol{S}\boldsymbol{V}^{l\dagger T}\boldsymbol{D}^{-1} \quad (39)$$

where the $l \times r$ Pascal-Vandermonde matrix $\boldsymbol{V}^{\boldsymbol{l}}$ and the $r \times r$ block-diagonal matrix $\boldsymbol{D}$ were introduced in Proposition 1, the r-dimensional vectors $\boldsymbol{v}'_{\downarrow}$ and $\boldsymbol{v}'_{\uparrow}$ are defined as

$$\boldsymbol{v}'_{\downarrow} = \frac{\boldsymbol{Z}^{-1}\boldsymbol{v}_{\downarrow}}{1 - \boldsymbol{v}_{\downarrow}^{H}\boldsymbol{Z}^{-1}\boldsymbol{v}_{\downarrow}}$$
$$\boldsymbol{v}'_{\uparrow} = \frac{\boldsymbol{Z}^{-1}\boldsymbol{v}_{\uparrow}}{1 - \boldsymbol{v}_{\uparrow}^{H}\boldsymbol{Z}^{-1}\boldsymbol{v}_{\uparrow}}$$

and the n-dimensional vectors $\boldsymbol{e}_{\downarrow}$ and $\boldsymbol{e}_{\uparrow}$ are defined as

$$\boldsymbol{e}_{\downarrow} = [0 \ldots 0, 1]^{T} - \boldsymbol{V}^{n}\boldsymbol{Z}^{-1}\boldsymbol{v}_{\downarrow}$$
$$\boldsymbol{e}_{\uparrow} = [1, 0 \ldots 0]^{T} - \boldsymbol{V}^{n}\boldsymbol{Z}^{-1}\boldsymbol{v}_{\uparrow}.$$

Corollary 11 is derived from proposition 10 by means of the basis change (7).

3) Perturbation of the Poles: Finally, we focus in this section on the perturbation of the poles. Theoretically, they are obtained by computing the Jordan form of the spectral matrix $\boldsymbol{\Phi}$. In practice, contrary to $\boldsymbol{\Phi}$, the perturbed spectral matrix does not have multiple eigenvalues: multiple poles are scattered into several single eigenvalues.

More precisely, it was shown in [2] that if $M_k > 1$:

- the first-order perturbation of the pole z_k is homogeneous and isotropic, so that the M_k perturbed eigenvalues $z_k^{(m)}(\varepsilon)$ form the vertices of a regular polynomial of order M_k in the complex plan;
- the perturbation of the scattered eigenvalues is of order ε^{1/M_k}, which suggests that multiple poles are more sensitive to perturbations than single poles.

In fact it is possible to overcome this problem by no longer considering the eigenvalues $z_k^{(m)}(\varepsilon)$ as M_k distinct estimators of the same pole z_k, but rather by forming a single estimator of this pole by averaging the $z_k^{(m)}(\varepsilon)$. The following proposition shows that the arithmetic mean $z_k(\varepsilon)$ of the scattered eigenvalues admits a series expansion.

Proposition 12 (Perturbation of the Poles): Let $z_k(\varepsilon) = (1/M_k)\sum_{m=0}^{M_k-1} z_k^{(m)}(\varepsilon)\ \forall k \in \{0 \ldots K-1\}$. Then the function $\varepsilon \mapsto z_k(\varepsilon)$ is $\mathcal{C}^{\infty}$ and admits the first-order expansion

$$z_k(\varepsilon) = z_k + \varepsilon\Delta z_k + O(\varepsilon^2) \quad (40)$$

where Δz_k is the complex number

$$\Delta z_k = \frac{1}{M_k}\mathrm{trace}\left(\Delta \boldsymbol{J}_k^{\perp}\right). \quad (41)$$

Here, $\Delta \boldsymbol{J}_k^{\perp}$ is the sub-block of dimension $M_k \times M_k$ extracted from the matrix $\Delta \boldsymbol{J}^{\perp}$, which is associated to z_k.[20] The complex number Δz_k can also be written in the form

$$\Delta z_k = \frac{1}{M_k\alpha_k^{(M_k-1)}}\left(\boldsymbol{e}_{\downarrow}{}^{H}\Delta\boldsymbol{S}\boldsymbol{f}_{\downarrow k} - \boldsymbol{e}_{\uparrow}{}^{H}\Delta\boldsymbol{S}\boldsymbol{f}_{\uparrow k}\right). \quad (42)$$

[20]This corresponds to the rows and columns of indices $\sum_{k'=0}^{k-1} M_{k'}$ to $\sum_{k'=0}^{k} M_{k'} - 1$.

The l-dimensional vectors $\boldsymbol{f}_{\downarrow k}$ and $\boldsymbol{f}_{\uparrow k}$ are defined[21] as

$$\begin{cases} \boldsymbol{f}_{\downarrow k} = \alpha_k^{(M_k-1)}\boldsymbol{V}^{l\dagger T}\boldsymbol{D}^{-1}\boldsymbol{J}\boldsymbol{v}'_{\downarrow k} \\ \boldsymbol{f}_{\uparrow k} = \alpha_k^{(M_k-1)}\boldsymbol{V}^{l\dagger T}\boldsymbol{D}^{-1}\boldsymbol{J}\boldsymbol{v}'_{\uparrow k} \end{cases} \quad (43)$$

where $\boldsymbol{v}'_{\downarrow k}$ and $\boldsymbol{v}'_{\uparrow k}$ are the r-dimensional vectors whose coefficients are equal to those of $\boldsymbol{v}'_{\downarrow}$ and $\boldsymbol{v}'_{\uparrow}$ inside the kth sub-block,[22] and zero outside this sub-block.

Equation (42) is derived by substituting (39) into (41). It can be noticed that the antihermitian matrix $\boldsymbol{A}$, introduced in proposition 9, is no longer involved in proposition 12. We can conclude that the performance of the generalized ESPRIT algorithm is not sensitive to the particular choice of the orthonormal basis $\boldsymbol{W}(\varepsilon)$ (at the first order).

Since the matrix $\Delta\boldsymbol{S}$ is Hankel and contains the samples of the PACE signal, the right member of (42) contains linear combinations of $\Delta s(\tau)$ for $\tau \in \{t-l+1 \ldots t+n-1\}$. Therefore, (42) can also be written as a scalar product

$$\Delta z_k = \frac{1}{M_k\alpha_k^{(M_k-1)}}\boldsymbol{u}_k{}^{H}\Delta\boldsymbol{s} \quad (44)$$

where for all $\tau \in [0, \ldots, n+l-2]$, the coefficient of index τ in $\boldsymbol{u}_k$ is[23]

$$u_k(\tau) = \sum_{\upsilon=0}^{n-1}\left(e_{\downarrow}(\upsilon)f_{\downarrow k}^{*}(\tau-\upsilon) - e_{\uparrow}(\upsilon)f_{\uparrow k}^{*}(\tau-\upsilon)\right)\mathbf{1}_{\{\tau-l<\upsilon\leq\tau\}}. \quad (45)$$

Equation (45) involves two convolution products, which are due to the Hankel matrix/vector products in (42).

Finally, proposition 3 in Section III shows how the perturbation of the poles affects the frequencies and damping factors.

B. Perturbation of the Amplitudes and Phases

We successively derive the following first-order expansions:

- $\boldsymbol{V}^{N}(\varepsilon) = \boldsymbol{V}^{N} + \varepsilon\Delta\boldsymbol{V}^{N} + O(\varepsilon^2)$, where the expression of $\Delta\boldsymbol{V}^{N}$ is a linear function of the Δz_k's (lemma 13);
- $\boldsymbol{\alpha}(\varepsilon) = \boldsymbol{\alpha} + \varepsilon\Delta\boldsymbol{\alpha} + O(\varepsilon^2)$, where the expression of $\Delta\boldsymbol{\alpha}$ is a linear function of Δs (proposition 14);
- $\begin{cases} a_k(\varepsilon) = a_k + \varepsilon\Delta a_k + O(\varepsilon^2) \\ \phi_k(\varepsilon) = \phi_k + \varepsilon\Delta\phi_k + O(\varepsilon^2) \end{cases}$ where the expressions of Δa_k and $\Delta\phi_k$ are functions of $\Delta\boldsymbol{\alpha}$ (corollary 4).

1) Perturbation of the Pascal-Vandermonde Matrix:

Lemma 13 (Perturbation of the Pascal-Vandermonde Matrix): Let $\boldsymbol{V}^{N}(\varepsilon)$ be the $N \times r$ Pascal-Vandermonde matrix associated to the estimated poles $\{z_0(\varepsilon), \ldots, z_{K-1}(\varepsilon)\}$ defined in proposition 12. Then the function $\varepsilon \mapsto \boldsymbol{V}^{N}(\varepsilon)$ is $\mathcal{C}^{\infty}$ in the neighborhood of $\varepsilon = 0$, and admits the first-order expansion

$$\boldsymbol{V}^{N}(\varepsilon) = \boldsymbol{V}^{N} + \varepsilon\Delta\boldsymbol{V}^{N} + O(\varepsilon^2) \quad (46)$$

[21]It can be verified that if $M_k = 1$, the vectors $\boldsymbol{f}_{\downarrow k}$ and $\boldsymbol{f}_{\uparrow k}$ do not depend on any complex amplitude.

[22]This corresponds to the coefficients of indices $\sum_{k'=0}^{k-1} M_{k'}$ to $\sum_{k'=0}^{k} M_{k'} - 1$.

[23]Here the function $\mathbf{1}_{(.)}$ is one if its argument is true and zero otherwise.

where the matrix $\Delta \boldsymbol{V}^N$ can be written in the form

$$\Delta \boldsymbol{V}^N = \overline{\boldsymbol{V}}^N \Delta \boldsymbol{Z} \tag{47}$$

where $\overline{\boldsymbol{V}}^N$ is the $N \times (r+K)$ Pascal-Vandermonde matrix obtained by concatenating the generalized Pascal matrices $\overline{\boldsymbol{C}}_k^N = \boldsymbol{C}_{M_k+1}^N(z_k)$ of dimension $N \times (M_k+1)$, and

$$\Delta \boldsymbol{Z} = \mathrm{diag}(\Delta \boldsymbol{Z}_0, \ldots, \Delta \boldsymbol{Z}_{K-1}) \tag{48}$$

is a $(r+K) \times r$ matrix whose diagonal blocks

$$\Delta \boldsymbol{Z}_k = \Delta z_k \times \begin{bmatrix} \boldsymbol{0}_{(1 \times M_k)} \\ \mathrm{diag}(1, 2, \ldots, M_k) \end{bmatrix} \tag{49}$$

have dimension $(M_k+1) \times M_k$.

Lemma 13 is proved by substituting the first-order expansion of the poles (40) into the expression of the Pascal-Vandermonde matrix, and by extracting the terms of order 1.

2) Perturbation of the Amplitudes and Phases: The perturbation of the complex-valued amplitudes can be derived from lemma 13. Let

$$\boldsymbol{s} = [s(t-l+1), \ldots, s(t+n-1)]^T$$

be the N-dimensional vector containing the samples of the PACE signal. For all $\varepsilon \in \mathbb{R}$, the N-dimensional observed vector is $\boldsymbol{s}(\varepsilon) = \boldsymbol{s} + \varepsilon \Delta \boldsymbol{s}$.

Proposition 14 (Perturbation of the Complex-Valued Amplitudes): The perturbed LS estimate defined in (9) can be written in the form

$$\boldsymbol{\alpha}(\varepsilon) = \boldsymbol{V}^N(\varepsilon)^\dagger \boldsymbol{s}(\varepsilon). \tag{50}$$

Then the function $\varepsilon \mapsto \boldsymbol{\alpha}(\varepsilon)$ is $\mathcal{C}^\infty$ in the neighborhood of $\varepsilon = 0$, and admits the first-order expansion

$$\boldsymbol{\alpha}(\varepsilon) = \boldsymbol{\alpha} + \varepsilon \Delta \boldsymbol{\alpha} + O(\varepsilon^2). \tag{51}$$

The r-dimensional vector $\Delta \boldsymbol{\alpha}$ satisfies

$$\Delta \boldsymbol{\alpha} = \boldsymbol{B}^H \Delta \boldsymbol{s} \tag{52}$$

where the $r \times N$ matrix

$$\boldsymbol{B}^H = \boldsymbol{V}^{N\dagger} \left(\boldsymbol{I}_N - \overline{\boldsymbol{V}}^N \begin{bmatrix} \boldsymbol{B}_0 \\ \vdots \\ \boldsymbol{B}_{K-1} \end{bmatrix} \right) \tag{53}$$

involves the $(M_k+1) \times N$ matrices $\boldsymbol{B}_k$, of rank 1

$$\boldsymbol{B}_k = \left[0, \frac{1}{M_k} \frac{\alpha_k^{(0)}}{\alpha_k^{(M_k-1)}}, \ldots, \frac{M_k-1}{M_k} \frac{\alpha_k^{(M_k-2)}}{\alpha_k^{(M_k-1)}}, 1 \right]^T \boldsymbol{u}_k^H. \tag{54}$$

Equations (51) –(54) are obtained by substituting (46) to (49) into the first-order expansion of (50).

Finally, proposition 4 in Section III shows how the perturbation of the complex-valued amplitudes influences the real-valued amplitudes and phases. The vector $\boldsymbol{b}_k^{(m)}$ is the column of $\boldsymbol{B}$ associated to the pole z_k at index m, i.e., the column of index $m + \sum_{k'=0}^{k-1} M_{k'}$. The derivation of (14) from (52) is straightforward.

REFERENCES

[1] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT—A subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1340–1342, Oct. 1986.

[2] R. Badeau, B. David, and G. Richard, "High resolution spectral analysis of mixtures of complex exponentials modulated by polynomials," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1341–1350, Apr. 2006.

[3] R. A. Serway and J. W. Jewett, *Physics for Scientists and Engineers*. Pacific Grove, CA: Brooks/Cole, 2003.

[4] B. Fischer and A. Medvedev, " $\mathbf{L}^2$ time delay estimation by means of laguerre functions," presented at the Amer. Control Conf., San Diego, CA, Jun. 1999.

[5] A. M. Sabatini, "Correlation receivers using laguerre filter banks for modelling narrowband ultrasonic echoes and estimating their time-of-flights," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 44, no. 6, pp. 1253–1263, Nov. 1997.

[6] O. V. Ivanova, L. Marcu, and M. C. K. Khoo, "A nonparametric method for analysis of fluorescence emission in combined time and wavelength dimensions," *Ann. Biomed. Eng.*, vol. 33, no. 4, pp. 531–544, Apr. 2005.

[7] T. W. Hänsch, A. L. Schawlow, and G. W. Series, "The spectrum of atomic hydrogen," *Sci. Amer.*, vol. 240, no. 94, pp. 531–544, Mar. 1979.

[8] P. W. Milonni and J. H. Eberly, *Lasers*, ser. Wiley Series in Pure and Applied Optics. New York: Wiley-Interscience, 1988.

[9] D. Filipovic, "Exponential-polynomial families and the term structure of interest rates," *Bernoulli*, vol. 6, no. 6, pp. 1081–1107, 2000.

[10] V. Slivinskas, M. Radavicius, and V. Simonyte, "Cramér-Rao bound for the estimates of requencies and damping factors of quasipolynomials in noise," Instit. Technol., Uppsala Univ., Sweden, Rep. UPTEC 92022R, Feb. 1992, Tech. Rep..

[11] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[12] R. Badeau, B. David, and G. Richard, "Cramér-Rao bounds for multiple poles and coefficients of quasipolynomials in colored noise," *IEEE Trans. Signal Process.*, submitted for publication.

[13] A. Kot, S. Parthasarathy, D. Tufts, and R. Vaccaro, "The statistical performance of state-variable balancing and Prony's method in parameter estimation," in *Proc. ICASSP'87*, Apr. 1987, vol. 12, pp. 1549–1552.

[14] P. Stoica and A. Nehorai, "Study of the statistical performance of the pisarenko harmonic decomposition method," *Inst. Elect. Eng. Proc. Radar Signal Process.*, vol. 135, no. 2, pp. 161–168, Apr. 1988.

[15] G. M. Riche de Prony, "Essai expérimental et analytique sur les lois de la dilatabilité de fluides élastiques et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alcool à différentes températures," (in French) *Journal de l'école polytechnique*, vol. 1, no. 22, pp. 24–76, 1795.

[16] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. R. Astron. Soc.*, vol. 33, pp. 347–366, 1973.

[17] R. Kumaresan and D. W. Tufts, "Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-30, no. 6, pp. 833–840, Dec. 1982.

[18] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.

[19] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 5, pp. 814–824, May 1990.

[20] B. Porat and B. Friedlander, "On the accuracy of the Kumaresan-Tufts method for estimating complex damped exponentials," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 2, pp. 231–235, Feb. 1987.

[21] Y. Hua and T. K. Sarkar, "Perturbation analysis of TK method for harmonic retrieval problems," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 2, pp. 228–240, Feb. 1988.

[22] P. Stoica and T. Söderström, "Statistical analysis of MUSIC and subspace rotation estimates of sinusoidal frequencies," *IEEE Trans. Signal Process.*, vol. 39, no. 8, pp. 1836–1847, Aug. 1991.
[23] A. Eriksson, P. Stoica, and T. Soderstrom, "Second-order properties of MUSIC and ESPRIT estimates of sinusoidal frequencies in high SNR scenarios," *Inst. Elect. Eng. Proc. Radar, Sonar Navigat.*, vol. 140, no. 4, pp. 266–272, Aug. 1993.
[24] Y. Hua and T. K. Sarkar, "On SVD for estimating generalized eigenvalues of singular matrix pencil in noise," *IEEE Trans. Signal Process.*, vol. 39, no. 4, pp. 892–900, Apr. 1991.
[25] S. Roman, *The Umbral Calculus*. New York: Academic, 1984, sec. 1.2: The Lower Factorial Polynomial.
[26] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, 2nd ed. Reading, MA: Addison-Wesley, 1994.
[27] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prenctice-Hall, 1980.
[28] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[29] R. Badeau, G. Richard, and B. David, "Sliding window adaptive SVD algorithms," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 1–10, Jan. 2004.
[30] R. Badeau, B. David, and G. Richard, "Fast approximated power iteration subspace tracking," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2931–2941, Aug. 2005.
[31] R. Badeau, B. David, and G. Richard, "Yet another subspace tracker," in *Proc. IEEE ICASSP'05*, Mar. 2005, vol. 4, pp. 329–332.
[32] H. Zeiger and A. McEwen, "Approximate linear realizations of given dimension via Ho's algorithm," *IEEE Trans. Autom. Control*, vol. AC-19, no. 2, p. 153, Apr. 1974.
[33] R. Badeau, B. David, and G. Richard, "A new perturbation analysis for signal enumeration in rotational invariance techniques," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 450–458, Feb. 2006.
[34] Z. Zeng, "On Ill-conditioned eigenvalues, multiple roots of polynomials, and their accurate computations," Northeastern Illinois University, Chicago, IL, MSRI Preprint No. 1998-048, 1998, Tech. Rep..
[35] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 2, pp. 387–392, Apr. 1985.
[36] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A subspace algorithm for certain blind identification problems," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 499–511, Mar. 1997.
[37] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Process.*, vol. 44, no. 1, pp. 95–107, Jan. 1995.
[38] R. Badeau, "Méthodes à haute résolution pour l'estimation et le suivi de sinusoïdes modulées. Application aux signaux de musique" Transl.: "High resolution methods for estimating and tracking modulated sinusoids. Application to music signals," Ph.D. dissertation, Télécom Paris (ENST), Paris, France, Apr. 2005 [Online]. Available: http://pastel.paristech.org/1234/
[39] R. Badeau, G. Richard, and B. David, "Performance of ESPRIT for estimating mixtures of complex exponentials modulated by polynomials: Supporting document," Télécom Paris (ENST)-CNRS LTCI, Paris, France, Oct. 2007.

Roland Badeau (M'02) was born in Marseilles, France, on August 28, 1976. He received the State Engineering degree from the École Polytechnique, Palaiseau, France, in 1999, the State Engineering degree from the École Nationale Supérieure des Télécommunications (ENST), Paris, France, in 2001, the M.Sc. degree in applied mathematics from the École Normale Supérieure (ENS), Cachan in 2001, and the Ph.D. degree from the ENST in 2005 in the field of signal processing.

In 2001, he joined the Department of Signal and Image Processing, GET-Télécom Paris (ENST), as an Assistant Professor, where he became Associate Professor in 2005. His research interests include high resolution methods, adaptive subspace algorithms, audio signal processing, and music information retrieval.

Gaël Richard (M'02–SM'06) received the State Engineering degree from the École Nationale Supérieure des Télécommunications (ENST), Paris, France, in 1990, the Ph.D. degree from LIMSI-CNRS, University of Paris-XI, in 1994 in speech synthesis and the Habilitation à Diriger des Recherches degree from the University of Paris XI in September 2001.

He then spent two years with the CAIP Center, Rutgers University, Piscataway, NJ, in the speech processing group of Prof. J. Flanagan, where he explored innovative approaches for speech production. Between 1997 and 2001, he successively worked for Matra Nortel Communications, Bois d'Arcy, France, and for Philips Consumer Comunications, Montrouge, France. In particular, he was the project manager of several large-scale European projects in the field of audio and multimodal signal processing. In September 2001, he joined the Department of Signal and Image Processing, GET-Télécom Paris (ENST), where he is now full Professor in audio signal processing and Head of the Audio, Acoustics and Waves Research Group. He is a coauthor of more than 70 papers and inventor in a number of patents, he is also one of the experts of the European commission in the field of audio signal processing and man/machine interfaces.

Prof. Richard is a member of the EURASIP and an Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING.

Bertrand David (M'06) was born on March 12, 1967 in Paris, France. He received the M.Sc. degree from the University of Paris-Sud, in 1991, and the Agrégation, a competitive French examination for the recruitment of teachers, in the field of applied physics, from the École Normale Supérieure (ENS), Cachan, France. He received the Ph.D. degree from the University of Paris 6 in 1999, in the fields of musical acoustics and signal processing of musical signals.

He formerly taught in a graduate school in electrical engineering, computer science, and communication. He also carried out industrial projects aiming at embarking a low complexity sound synthesizer. Since September 2001, he has worked as an Associate Professor with the Signal and Image Processing Department, GET-Télécom Paris (ENST). His research interests include parametric methods for the analysis/synthesis of musical and mechanical signals, music information retrieval, and musical acoustics.