PREDICTION ERROR METHODS FOR TIME-DOMAIN BLIND IDENTIFICATION OF MULTICHANNEL FIR FILTERS.

K. Abed Meraim , P. Duhamel*, D. Gesbert**, P. Loubaton*, S. Mayrargue**, E. Moulines*, D.Slock ****

* Télécom Paris / URA 820 and GDR 134/CNRS 46, rue Barrault, 75634 Paris CEDEX 13 FRANCE
*** Eurecom Institute, 2229 route des Crêtes, BP.193, 06904 Sophia Antipolis Cedex, FRANCE
** France Télécom/CNET/PAB and GDR 134/CNRS 38-40 rue du General Leclerc, 92131 Issy-Les-Moulineaux, FRANCE

ABSTRACT

Blind channel identification methods based on the oversampled channel output is a problem of current theoretical and practical interest. In this contribution, it is first demonstrated that the subspace methods developped in [1] are not robust to errors in the determination of the model order. An alternative solution is then proposed, based on a linear prediction approach. The effect of overestimating the channel order is investigated by simulations: it is demonstrated that the prediction error method is "robust" to over-determination, in contrast to most of the schemes suggested to date.

1. INTRODUCTION

Since the early work by Sato [2], most of the blind identification/equalization techniques have been based on the use (implicit or not) of higher-order statistics, which are known to suffer from many drawbacks (in terms of finite-sample variance, robustness to noise...). The recent proposal by Gardner [3] and Tong, Xu and Kailath [4] of methods allowing the blind identification of the channels *using only second order statistics* appears as a major breakthrough in this field. The basic idea motivating these approaches consists in recognizing that appropriately oversampled communication signals are cyclostationary, and that, under a mild additional hypothesis, the phase of linear time-invariant channel can be retrieved from the periodically time-varying channel output correlation function [3, 4]. Since these algorithms rely on fractional sampling and second-order statistics, they show several desirable properties, making it suitable in mobile communication context (see, [4]): at the first place, these algorithms require fewer symbols than most of the schemes suggested to date, making these solutions attractive even in the presence of relatively rapid channel variation. Some improvements over these methods have been presented in [1] (subspace method) and [5] (deterministic maximum likelihood method). Our contribution in this paper is twofold. (i) The potential drawbacks of the subspace and related methods (proposed in [5, 1]), when the exact order of the system is unknown, are outlined and (more importantly) (ii) a time-domain procedure based on a prediction approach, which is more robust to the overdetermination of the system order is presented.

*The authors are listed in alphabetic order. This contribution is the result of a collective research effort carried out in the working groups GT4 -multivariate systems- and GT8 -communication and coding- inside the GDR 134-CNRS.

2. SUBSPACE METHODS

The continuous-time output from a linear time-invariant channel driven by a PAM/QAM sequence is given by

$$\tilde{y}(t) = \sum_k s(k)h(t-kT) + \tilde{w}(t) \qquad (1)$$

where $s(k)$ is an independent and identically distributed (i.i.d) sequence of symbols ($Es(k) = 0$, $E(s(k)^2) = 1$), $\tilde{w}(t)$ is an additive temporally white noise $E(\tilde{w}(t_1)\tilde{w}(t_2)) = \sigma^2\delta(t_1 - t_2)$, T is the baud rate, and $h(t)$ is the impulse response of the cascade of channel, transmit and receive filters, assumed to be causal and time-limited to $(M+1)$ symbols duration. Oversampling by a factor $q > 1$ leads to a $q \times 1$ discrete sequence $y(n) = [y_1(n), \cdots, y_q(n)]^T$ (respectively $w(n) = [w_1(n), \cdots, w_q(n)]^T$), where $y_l(n) = \tilde{y}((l-1)T/q + nT + t_0)$ (respectively $w_l(n) = \tilde{w}((l-1)T/q + nT + t_0)$). According to (1), $y(n)$ can be seen as the noisy output of a $q \times 1$ time-invariant polynomial transfer function $\mathrm{h}(z)$ driven by the scalar sequence $s(n)$, i.e.

$$y(n) = [\mathrm{h}_1(z), \cdots, \mathrm{h}_q(z)]^T s(n) + w(n) = [\mathrm{h}(z)]s(n) + w(n)$$

$$\mathrm{h}_l(z) = \sum_{k=0}^{M} h((l-1)\frac{T}{q} + kT + t_0)z^{-k} = \sum_{k=0}^{M} h_l(k)z^{-k}$$

Denote by $r(\tau)$ the autocovariance function of $y(n)$, $r(\tau) = E\{y(n+\tau)y(n)^T\}$. Under the fundamental hypothesis $\mathrm{h}(z) \neq 0$ for *each* z, it has been shown in [4] that $\mathrm{h}(z)$ is identifiable from a finite number of values of $r(\tau)$. A subspace-based identification scheme was later introduced in [1]. It relies on geometric properties of the so-called Sylvester $q(N+1) \times (M+N+1)$ block-Toeplitz matrix associated with the polynomial vector $\mathrm{h}(z)$, defined as:

$$\mathcal{T}_N(h) = \begin{pmatrix} h(0) & \cdots & h(M) & 0 & \cdots & 0 \\ 0 & h(0) & \cdots & h(M) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & h(0) & \cdots & h(M) \end{pmatrix}$$

Denoting by $\mathcal{R}_N$ the covariance matrix of the $q(N+1) \times 1$ vector $Y_N(n) = [y(n), \cdots, y(n-N)]^T$, it comes, by using

0-7803-2431-5/95 $4.00 © 1995 IEEE

the signal model (2) and the noise properties:

$$\begin{aligned} Y_N(n) &= \mathcal{T}_N(h)s_{M+N}(n) + W_N(n) \\ s_{M+N}(n) &= [s(n), \cdots, s(n-M-N)]^T \\ \mathcal{R}_N &= \mathcal{T}_N(h)\mathcal{T}_N^T(h) + \sigma^2 I \end{aligned}$$

Since $q > 1$, $\mathcal{R}_N - \sigma^2 I$ is singular as soon as $q(N+1) > (N+1+M)$. Denote Π the orthogonal projection matrix onto the noise subspace of $\mathcal{R}_N$ (i.e the orthogonal subspace of Range($\mathcal{T}_N(h)$)). The subspace identification method is ultimately related to the following theorem:

Theorem 1 *Assume that* **(H1)** $h(z) \neq 0 \ \ \forall z$, *then*

1. *for $N \geq M-1$, the matrix $\mathcal{T}_N(h)$ is full rank.*
2. *let* $\mathrm{f}(z)$ *be a* $q \times 1$ *polynomial transfer function,* $\deg(\mathrm{f}(z)) = P$. *For $N \geq M$, we have*

$$\Pi\mathcal{T}_N(f) = 0 \Longleftrightarrow \mathrm{f}(z) = \mathrm{p}(z)\mathrm{h}(z) \qquad (2)$$

where $\mathrm{p}(z)$ *is some scalar polynomial. In this case, the degree P must be greater than M.*

In practice the orthogonal projector $\hat{\Pi}$ is estimated from the observed signals covariance matrix $\hat{\mathcal{R}}_N$, and, if M is known, h can be estimated (up to a constant) by minimizing on the set of all $q \times 1$ degree-M polynomials the following quadratic criterion

$$\mathrm{Trace}(\mathcal{T}_N^T(f)\hat{\Pi}\mathcal{T}_N(f)) \qquad (3)$$

under a suitable constraint (see [1] for more details). If M is overestimated, the argument $\hat{\mathrm{f}}(z)$ of (3) will represent an estimate of a certain polynomial $r(z)\mathrm{h}(z)$, and another algorithm would here be required to factorize from the components of $\hat{\mathrm{f}}(z)$ some approximatively common zeros. But, using such an algorithm would require some test to decide how many zeros are common, and the relevance of the final estimate should again depend on the result of the test. The same conclusion holds for the method proposed in [5], which is closely related to the above subspace method. The method proposed in [4] estimates the Sylvester matrix $\mathcal{T}_N(h)$ without using the specific parametrisation of the matrix. Based on simulations, it seems to be 'more' robust to the errors in the estimation of the model order than subspace methods, but, as it will be shown in section 4, its performance is still somewhat less inferior in comparison with those of the linear prediction method presented in the following.

3. PREDICTION BASED METHODS

In this section, a linear prediction approach is developped to identify $\mathrm{h}(z)$. This approach has been first investigated by Slock in [6] in the special case where $q = 2$, and when the degree M of $\mathrm{h}(z)$ is known. The main contribution of this section is to generalize the results of [6], and more importantly, to show that the linear prediction approach is robust to an over-determination of the order M. In order to simplify the exposition, the noiseless case $w(n) = 0$ is considered in a first part. A straightfroward extension to noisy data is presented in a second part.

The basic idea behind the linear prediction approach is to recognize that the moving-average (MA) process $y(n) = [\mathrm{h}(z)]v(n)$ is also a *finite order* autoregressive process (AR). This property is related to the generalized Bezout identity (see, for example, [7]). Under **(H2)** ($\mathrm{h}(z)$ is irreducible), it is known [7] that there exists a $1 \times q$ polynomial vector $\mathrm{g}(z)$ such that $\mathrm{g}(z)\mathrm{h}(z) = 1$. By applying $\mathrm{g}(z)$ to $y(n)$, it comes that $[\mathrm{g}(z)]y(n) = s(n)$: $h(z)$ can be *exactly* inverted by an FIR causal filter. This relation is the key behind all subsequent derivations.

In order to proceed, some additional notations and definitions are in order. Denote $H_{n-1}(y)$ the past of y up to time $n-1$ defined by

$$H_{n-1}(y) = \mathrm{sp}\{y_i(n-l)/\, i = 1, q,\ l \geq 1\} \qquad (4)$$

Here, $\mathrm{sp}\{x_i \in I\}$ stands for the Hilbert subspace generated by $\{x_i \in I\}$. Denote accordingly $H_{n-1,N}(y)$ the finite past of y,

$$H_{n-1,N}(y) = \mathrm{sp}\{y_i(n-l)/\, i = 1, q,\ 1 \leq l \leq N\} \qquad (5)$$

The innovation process $(i(n))_{n\in\mathbb{Z}}$ of y is the q–variate white noise sequence defined by

$$i(n) = y(n) - y(n)|_{H_{n-1}(y)} \qquad (6)$$

where $|$ stands for the orthogonal projection in H. The process y will be said to be autoregressive of order N if $i(n)$ coincides with the finite order innovation sequence $i_N(n) = y(n) - y(n)|_{H_{n-1,N}(y)}$. We have the following theorem.

Theorem 2 *Under* **(H1-2)**, $y(n)$ *is an order M autoregressive process. Its innovation process is given by* $i(n) = h(0)v(n)$.

proof: Consider $\mathcal{T}_N(\mathrm{h})$ the Sylvester matrix associated to the filter $\mathrm{h}(z)$. Under **(H1)**, for $N = (M-1)$, $\mathcal{T}_N(\mathrm{h})$ is full rank, and is thus left invertible, i.e. it exists a $(2M \times qM)$ matrix K such that $K\mathcal{T}_{M-1}(\mathrm{h}) = I$. Therefore, $H_{n,M-1}(y) = H_{n,2M-1}(s)$, so that $H_n(y) = H_n(s)$, the infinite past of processes y and s coincide. Since $y(n) = h(0)s(n) + \sum_{k=1}^{M} h(k)s(n-k)$, $s(n)$ is a white noise, and $-\sum_{k=1}^{M} h(k)s(n-k) = y(n)|_{H_{n-1}(y)}$, the innovation of $y(n)$ is given by: $i(n) = h(0)s(n)$. As $H_{n,M-1}(y) = H_{n,2M-1}(s)$, it holds $H_{n-1,M}(y) = H_{n-1,2M}(s)$. Thus, the variables $s(n-k), 1 \leq k \leq M$ belong to $H_{n-1,M}(y)$, and hence,

$$-\sum_{k=1}^{M} h(k)s(n-k) = y(n)|_{H_{n-1}(y)} \in H_{n-1,M}(y)$$

In some sense, $s(n)$ represents the normalized innovation of $y(n)$, and a $1 \times q$ polynomial filter $\mathrm{g}(z)$ verifying $[\mathrm{g}(z)]y(n) = s(n)$ can be seen as a *prediction error filter* (PEF). Such a filter can be easily computed by using generalized Yule-Walker equations, as shown below. The important point is that the polynomial transfer function $\mathrm{h}(z)$ can be recovered from any PEF $\mathrm{g}(z)$ by using $h(k) = E(y(n)([\mathrm{g}(z)]y(n-k))^T)$.

The innovation $i(n)$ is computed by projecting $y(n)$ onto the space generated by the random variables $\{y_i(n-l)/i = 1, q,\ l = 1, P\}$ where $P \geq M$. Let $[A(1), \cdots, A(P)]$ be $q \times q$ matrices such that $y(n) + \sum_{k=1}^{P} A(k)y(n-k) = i(n)$:

$$[A(1), \cdots, A(P)]\mathcal{R}_{P-1} = -[r(1), \cdots, r(P)] \qquad (7)$$

Since the innovation is not full-rank, these coefficients are not uniquely defined. A particular set of coefficients may be obtained by:

$$[A(1),\cdots,A(P)] = -[r(1),\cdots,r(P)]\mathcal{R}^{\#}_{P-1} \qquad (8)$$

where $\mathcal{R}_{P-1}$ is the covariance matrix of $Y_{P-1}(n) = [y(n),\cdots,y(n-P+1)]^T$ and the supersript # denotes the Moore-Penrose pseudo-inverse. According to theorem 2, the covariance of the innovation $D = r(0) + \sum_{k=1}^{P} A(k)r^T(k)$ is of rank one; the non-zero eigenvalue of D is equal to $\lambda_d = \|h_0\|^2$; the associated eigenvector d is the unit-norm $d = \pm h_0/\|h_0\|$ (the sign is not identifiable at the second-order). Take $l = d/\sqrt{\lambda_d}$. It comes

Lemma 1 $g(z) = l^T(I + \sum_{k=1}^{P} A(k)z^{-k})$ *is a* $1 \times q$ *polynomial vector such that* $[g(z)]y(n) = s(n)$.

Comments

- For the prediction method, when $P > M$, $h(M+1),\cdots,h(P)$ are equal to zero. And thus, when exact statistics are available, an exact identification of the filter coefficients is achieved under the solely hypothesis that the prediction order P is greater than the true model order. In practice, when only sample-estimate statistics are available the order estimation is necessary to perform the pseudo-inverse of the covariance matrix $\mathcal{R}_{P-1}$. However, a failure of the order determination procedure, doesn't affect seriously the performances of the method. Heuristically, the reasons are twofolds. First, let $P > M$ and let u a 'noise' eigenvector of matrix $\mathcal{R}_{P-1}$: $\mathcal{R}_{P-1}u = 0$; from (7), it comes that $[r(1),\ldots,r(P)]u = 0$, the noise eigenvectors are orthogonal to $[r(1),\cdots,r(P)]^T$. This relation is exact for ensemble-averaged covariance; it is expected to be approximately verified for sample-statistics. Next, the order determination method will fail when sample noise eigenvalues are closed to some signal eigenvalues (in the case where there is a clear cut between the two sets, the correct model order can be inferred directly). Then, it is likely that the $\hat{M} - M$ dominant noise eigenvalues are bounded away from zero. This two reasons together imply that the inversion does not cause serious numerical troubles. All these claims are supported by the numerous numerical simulations we have conducted, as illustrated in section 4.
- The prediction method can be straightforwardly extended to deal with noisy signals. In this case, it suffices to estimate σ^2 and to deal with the noise-free covariance, obtained by substracting to the observed signal covariance the noise covariance $\sigma^2 I$.

A practical implementation From the results of the previous section, and assuming the additive noise spatially and temporally white, we derive the following algorithm.

1. Given $T > P$ samples of the process y, $y_T = [y(1),\cdots,y(T)]$, compute $(P+1)$ (P is greater than M_e the order of the filter) autocovariance coefficients according to

$$\begin{aligned}\hat{r}(n) &= [\hat{r}_{ab}(n)]_{1\le a,b\le q} \\ &= \frac{1}{T-n}\sum_{t=1}^{T-n} y(t+n)y(t)^T \quad 0 \le n \le P\end{aligned}$$

From these coefficients, a sample estimate $\hat{\mathcal{R}}_P$ of the $q(P+1)\times q(P+1)$ covariance matrix of the vectorized observation $Y_P(n)$ can be obtained using the following construction

$$\hat{\mathcal{R}}_P = \begin{pmatrix} \hat{r}(0) & \hat{r}(1) & .. & \hat{r}(P) \\ \hat{r}(1)^T & \hat{r}(0) & .. & \hat{r}(P-1) \\ \vdots & \vdots & & \vdots \\ \hat{r}(P-1)^T & .. & \hat{r}(0) & \hat{r}(1) \\ \hat{r}(P)^T & .. & \hat{r}(1)^T & \hat{r}(0) \end{pmatrix}$$

2. From the eigen decomposition of $\hat{\mathcal{R}}_P$, estimate (i) the order $\hat{M}$ of the filter $h(z)$ by using, for example a standard information criteria test, such as the MDL test (see [8] and the references therein), (ii) the noise covariance $\hat{\sigma}^2$ as the average of the $N = q(P+1)-(\hat{M}+P+1)$ smallest eigenvalues of $\hat{\mathcal{R}}_P$.
3. Compute the vectorial prediction error filter $\hat{A}$ as

$$\begin{aligned}\hat{A} &= [I, \hat{A}(1),\ldots,\hat{A}(\hat{M})]^T \\ [\hat{A}(1),\ldots,\hat{A}(\hat{M})] &= -[\hat{r}(1),\ldots,\hat{r}(\hat{M})](\hat{\mathcal{R}}_{(\hat{M}-1)} - \hat{\sigma}^2 I)^{\#}\end{aligned}$$

4. Estimate the filter coefficient $l = h(0)/\parallel h(0)\parallel^2$ from the eigen-decomposition of the estimated covariance matrix $\hat{D}$:

$$\begin{aligned}\hat{D} &= [\hat{A}(1),\ldots,\hat{A}(\hat{M})][\hat{r}(1),\ldots,\hat{r}(\hat{M})]^T + \hat{r}(0) - \hat{\sigma}^2 I_q \\ \hat{D} &= \hat{V}\hat{\beta}\hat{V}^T \\ \hat{V} &= [\hat{v}_1,\ldots,\hat{v}_q], \quad \hat{V}^T\hat{V} = I_q \\ \hat{\beta} &= diag(\hat{\beta}_1,\ldots,\hat{\beta}_q), \quad \hat{\beta}_1 \ge \hat{\beta}_2 \ge \ldots \ge \hat{\beta}_q \\ \hat{l} &= \frac{\hat{v}_1}{\sqrt{\hat{\beta}_1}}\end{aligned}$$

5. Compute the $1 \times q(\hat{M}+1)$ prediction error filter $\hat{g}$ as:

$$\hat{g} = \hat{l}^T\hat{A}$$

6. If $\hat{S}(\hat{M})$ is the $q(\hat{M}+1)\times q(\hat{M}+1)$ matrix

$$\hat{S}(\hat{M}) = \begin{pmatrix} \hat{r}(\hat{M}) & 0 & 0 & \cdots & 0 \\ \hat{r}(\hat{M}-1) & \hat{r}(\hat{M}) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{r}(0)-\hat{\sigma}^2 I_q & \hat{r}(1) & \cdots & \cdots & \hat{r}(\hat{M}) \end{pmatrix}$$

estimate the $q(\hat{M}+1)\times 1$ filter coefficient vector $\hat{h} = [\hat{h}(\hat{M})^T,\ldots,\hat{h}(0)^T]^T$ as: $\hat{h} = \hat{S}(\hat{M})\hat{g}^T$

4. SIMULATION RESULTS

We present here some numerical simulations to assess the performance of our algorithm. We run simulations for a 4-variate ($q = 4$) MA model, with coefficients given by ($M = 5$): $h_1(z) = -0.04 - 0.30z^{-1} - 1.28z^{-2} - 0.53z^{-3} + 0.14z^{-4} - 0.26z^{-5}$, $h_2(z) = 0.91 - 0.20z^{-1} - 0.44z^{-2} - 1.02z^{-3} - 0.54z^{-4} - 0.08z^{-5}$, $h_3(z) = -1.18 + 0.49z^{-1} - 0.31z^{-2} + 0.40z^{-3} + 0.13z^{-4} - 1.85z^{-5}$ and $h_4(z) = 1.30 + 0.05z^{-1} + 0.34z^{-2} - 0.03z^{-3} + 0.40z^{-4} + 0.88z^{-5}$. The output observation noise is an i.i.d. sequence of zero-mean Gaussian variables and the input signal is an i.i.d sequence of zero-mean, unit-variance Gaussian variables independent from the noise signal.

The number of samples T is held constant ($T = 100$). For comparison, the algorithm of [4] was also simulated for the same channels. The criterion used in this section is the mean-square estimation error (MSE) defined as the square-root of the sample average, over the Monte-Carlo simulations, of the total estimation variance. For each experiment, 300 independent Monte-Carlo simulations are performed.

In figure 1, the *a priori* estimated order is set to $P = 7$, the input signal is Gaussian and the noise variance is varied between -30 dB and 0 dB. The AIC model-order determination procedure is used for both methods. The figure shows, the MSE in dB for the linear prediction method (respectively, the method of [4]) in solid line (respectively, in dashed line) as a function of the noise power. This show that our method consistently provides more reliable estimates is the case of unknown model-order M.

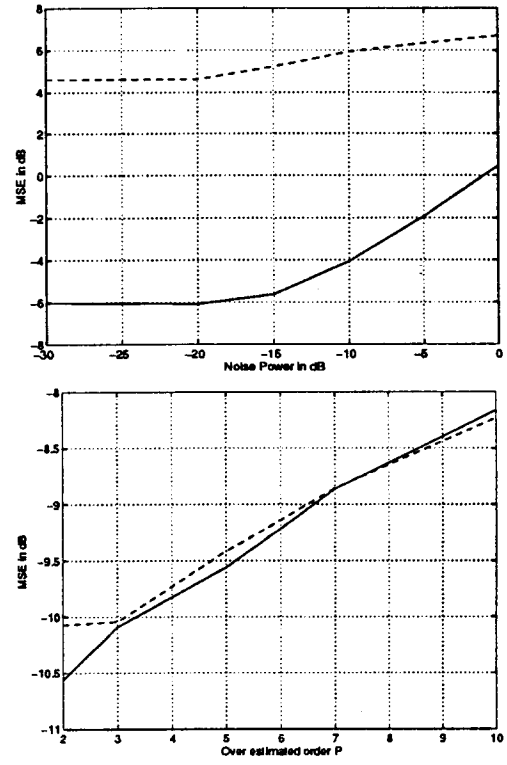
In the second experiment, we consider a two-variate ($q = 2$) MA model of order $M = 2$. The noise variance is set to -10 dB, the number of samples T is held constant ($T = 1000$) and the estimated order P is varied between 2 (i.e the exact order) and 10. To deal with a difficult order estimation context, the transfert functions are choosen as $h_1(z) = 1.1650 + 0.0751z^{-1} - 0.6965z^{-2}$ and $h_2(z) = 0.0352 - 0.0697z^{-1} + 0.1696z^{-2}$. The eigenvalues of the exact noise-free covariance matrix are given by {0, 0.0152, 0.0510, 1.0464, 1.8581, 2.6775}.

Fig.2 shows the MSE of the total estimation variance of $\hat{h} = [\hat{h}(0)^T, \ldots, \hat{h}(P)^T]^T$ of the linear prediction method when the MDL procedure is used for order determination (dashed-line) and this of the linear prediction method when the exact order is *a priori* known (solid-line), as a function of the over-estimated filter degree P. The table 1 shows the error-rate (ER) of the order estimation. This shows the robustness of the method to the errors in the estimation of the model order.

5. CONCLUSION

This contribution presents a linear prediction approach to the blind identification of multichannel FIR filters. This technique allows to obtain a good estimation behaviours of the channel coefficients even in the case where the number of coefficients is over-estimated. Numerical simulations have been preformed to evidence the usefulness of the method and to support our theoretical claims.

M	2	3	5	7	10
ER	0.62	0.51	0.67	0.67	0.79



Tab1. Error-rate on the estimated order by the MDL criterion

REFERENCES

[1] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters." To appear in IEEE Trans. On Signal Proc., 1994.

[2] Y. Sato, "A method for self recovering equalization," *IEEE Tr. on Comm.*, vol. 23, no. 6, pp. 679–682, 1975.

[3] W. Gardner, "A new method of channel identification," *IEEE Tr. on Comm.*, vol. 39, pp. 813–817, August 1991.

[4] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *Proc. of the 25th Asilomar Conference, Pacific Grove, CA*, pp. 856–860, 1991.

[5] H. Liu, G. Xu, and L. Tong, "A deterministic approach to blind identification of multichannel FIR systems," in *Proc. Int. Conf. on Acoust. Speech and Sig. Proc.*, vol. 4, pp. 581–584, 1994.

[6] D. Slock, "Blind fractionally-spaced equalization, perfect-reconstruction filter-banks and multichannel linear prediction," in *Proc. Int. Conf. on Acoust. Speech and Sig. Proc.*, vol. 4, pp. 585–588, 1994.

[7] T. Kailath, *Linear systems.* Prentice-Hall, 1980.

[8] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Tr. on Sig. Proc.*, vol. 33, pp. 387–392, Apr. 1985.