ON THE STABILITY OF CONSTRAINED LINEAR PREDICTIVE MODELS

Tom Bäckström and Paavo Alku

Helsinki University of Technology (HUT), Laboratory of Acoustics and Audio Signal Processing
P.O.Box 3000, FIN-02015 HUT, email: Tom.Backstrom@hut.fi

ABSTRACT

Stability of the all-pole model in conventional, unconstrained linear prediction with the autocorrelation criterion is well known. By exerting constraints to the optimisation problem it is possible to define models of order $m+l$ with m parameters. However, traditionally constraints have led to models whose stability is not guaranteed. In this paper, we will discuss constrained linear predictive models where the constraint is one-dimensional ($l = 1$) and derive stability criteria for these models.

1. INTRODUCTION

Linear prediction (LP) is a classical method in digital signal processing. Applications of LP have appeared in a remarkably wide variety including such fields as speech processing, geology and economics [1]. In many cases, constraints derived from a priori information can be used to funnel the linear predictive model to a favourable direction or to avoid illegal solutions. However, in applications that require the minimum-phase property, that is, the stability of the all-pole model, few such models are available. On the other hand, analysis of the stability of polynomials has been studied in detail in, among others [2, 3].

In our earlier work, we have studied constrained linear predictive models for speech processing [4, 5, 6] as well as stable all-pole models with alternative optimisation procedures [7]. The current work presents stability criteria for constrained LP models.

2. BACKGROUND

2.1. Conventional Linear Prediction

The formulation of the conventional linear prediction as described in [1] can be presented in matrix notation as follows. Using signal x_n and parameters a_i ($0 \le i \le m$), denoted $\mathbf{x}$ and $\mathbf{a}$ in vector notation, respectively, and model order m, we define the residual as $e_{LP,n} = \mathbf{x}^T\mathbf{a}$. We then minimise the expected value of the squared error $E\left[e_{LP,n}^2\right] = E\left[\mathbf{a}^T\mathbf{x}\mathbf{x}^T\mathbf{a}\right] = \mathbf{a}^T\mathbf{R}\mathbf{a}$ subject to the constraint $a_0 = 1$ or equivalently $\mathbf{a}^T\mathbf{b} = 1$ with $\mathbf{b} = [1,\,0,\,0,\,\dots]^T$.

The standard method for inclusion of constraints in optimisation is through usage of the Lagrange multiplier λ. The objective function is then

$$\eta(\mathbf{a},\lambda) = \mathbf{a}^T\mathbf{R}\mathbf{a} + \lambda\left(\mathbf{a}^T\mathbf{b} - 1\right). \quad (1)$$

The extreme is found through differentiation $0 = \partial\eta/\partial\mathbf{a} = 2\mathbf{R}\mathbf{a} + \lambda\mathbf{b}$. Multiplying from the left hand side with $\mathbf{a}^T$ yields $\mathbf{a}^T\mathbf{R}\mathbf{a} = E[e_{LP,n}^2] = -\lambda/2 = \sigma^2$. The optimal solution can then be written in the extended normal equations (also known as the extended Yule-Walker equations) as

$$\mathbf{R}\mathbf{a} = \sigma^2\mathbf{b}, \quad (2)$$

where σ^2 is the energy of the prediction error.

2.2. Line Spectral Pair Polynomials

The Line Spectral Pair (LSP) polynomials are defined using the transfer function $A(z)$ corresponding to vector $\mathbf{a}$ as $P(z) = A(z) + z^{-m-1}A(z^{-1})$ and $Q(z) = A(z) - z^{-m-1}A(z^{-1})$ [8]. It is well known that polynomials $P(z)$ and $Q(z)$ have the intra-model interlacing property, that is, the zeros of $P(z)$ and $Q(z)$ will be interlaced on the unit circle if $A(z)$ is minimum-phase. Correspondingly, if the zeros of $P(z)$ and $Q(z)$ are separate and interlaced on the unit circle, then the reconstruction $A(z) = \frac{1}{2}\left[P(z) + Q(z)\right]$ will be minimum-phase.

In addition to the intra-model interlacing property, the polynomials $P(z)$ and $Q(z)$ have also an inter-model interlacing property [9]. Namely, the zeros of the order m polynomials $P_m(z)$ and $Q_m(z)$ are interlaced with the zeros of the order $m+1$ polynomials $P_{m+1}(z)$ and $Q_{m+1}(z)$, respectively.

Predictors $\mathbf{p}$ and $\mathbf{q}$ corresponding to the LSP polynomials can be calculated directly from a similar relation as Eq. 2 by defining $\mathbf{R}\mathbf{p} = \sigma_p^2[1,\,0,\,\dots,0,\,1]^T$ and $\mathbf{R}\mathbf{q} = \sigma_q^2[1,\,0,\,\dots,0,\,-1]^T$. These polynomials have trivial zeros at $z = \pm 1$ [10], and if these zeros are removed, then the corresponding polynomials, denoted here by $S(z)$ and $T(z)$, respectively, can be calculated from equations $\mathbf{R}\mathbf{s} = \sigma_s^2[1,\,1,\,\dots,1,\,1]^T$ and $\mathbf{R}\mathbf{t} = \sigma_t^2[1,\,-1,\,1,\,-1,\,\dots]^T$, respectively [6]. The inter-model interlacing property can then be interpreted as interlacing of zeros of $P(z)$, $Q(z)$, $S(z)$ and $T(z)$ (excluding $S(z)$ or $T(z)$ for m odd).

3. CONSTRAINED LINEAR PREDICTION

Consider a linear model that estimates a future sample of signal x_n from past values of a filtered signal $\tilde{x}_n = c_n * x_n$, where c_n is the impulse response of a causal FIR filter (with the trivial case $c_n \equiv 0$, $n \geq 0$ excluded). The prediction residual is $e_n = x_n + \sum_{i=0}^{m-1} a_i \tilde{x}_{n-i} = x_n + \sum_{i=0}^{m-1} a_i c_{n-i} * x_{n-i} = x_n + \sum_{i=0}^{m-1} \sum_{k=0}^{l} a_i c_k x_{n-i-k}$. This equation is, importantly, different from conventional linear prediction since it defines a predictor of order $m+l-1$ using m parameters. Hence, we will call this approach *constrained* linear prediction. The residual can be concisely written as

$$e_n = \mathbf{b}^T \mathbf{x} + \mathbf{a}^T \mathbf{C}^T \mathbf{x}, \tag{3}$$

where $\mathbf{x} = [x_0, \ldots, x_{m+l-1}]^T$, $\mathbf{a} = [a_0, \ldots, a_{m-1}]^T$ is the parameter vector and $\mathbf{b}_i = \delta_i$. Matrix $\mathbf{C} \in \mathbb{R}^{(m+l) \times m}$ is defined such that $\mathbf{C}_{ij} = c_{j-i}$ for $0 \leq j - i \leq l$, where $l+1$ is the length of c_n. Later, we will need the null-space of matrix $\mathbf{C}$, defined as $\mathbf{C}^T \mathbf{C}_0 = 0$ where $\mathbf{C}_0 \in \mathbb{R}^{(m+l) \times l}$.

For the simple, one parameter FIR filter with coefficient vector $[1, -c]^T$ (i.e. $l = 1$), the corresponding constraint matrix is

$$\mathbf{C}^T = \begin{bmatrix} 1 & -c & 0 & \cdots & 0 \\ 0 & 1 & -c & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -c \end{bmatrix} \tag{4}$$

and its null space is $\mathbf{C}_0 = \left[1, c^{-1}, c^{-2}, \ldots, c^{-(m+l-1)}\right]^T$. Matrix $\mathbf{C}_0$ will thus be equal to the LSP constraints (with one trivial zero removed) for $c = \pm 1$ [6].

Note that we have chosen to predict sample x_n from $\tilde{x}_{n-i}$ where i ranges from *zero* to $m-1$. The sample x_n is therefore estimated from the *current* and past samples of $\tilde{x}_n$. One would therefore easily be led to believe that the predictor is non-causal. Luckily, however, if the FIR filter is non-trivial then the residual e_n can be determined since its computation (Eq. 3) contains terms of x_{n-i} where $i \in [0, m+l-1]$, and the optimisation problem is unambiguous. The transfer function of predictors obtained by these formulations will therefore have a coefficient of z^0 that is generally not equal to one.

We can then formulate our optimisation problem as:

Problem Minimise $E[e_n^2]$
with $e_n = \mathbf{b}^T \mathbf{x} + \mathbf{a}^T \mathbf{C}^T \mathbf{x}$

Solution We have the auxiliary equation $\mathbf{R}\mathbf{h} = \mathbf{C}_0 \lambda$ with the unknown vector $\mathbf{h}$. The $l \times 1$ vector λ is solved from $\mathbf{C}_0^T \mathbf{R}^{-1} \mathbf{C}_0 \lambda = \mathbf{C}_0^T \mathbf{b}$. Vector $\mathbf{a}$ can be calculated from the overdetermined equation $\mathbf{h} - \mathbf{b} = \mathbf{C}\mathbf{a}$.

Proof Defining $\mathbf{h} = \mathbf{b} + \mathbf{C}\mathbf{a}$ yields $e_n = \mathbf{x}^T \mathbf{h}$. Obviously, $\mathbf{h} - \mathbf{b} = \mathbf{C}\mathbf{a}$ and $\mathbf{h} - \mathbf{b}$ is thus in column space of $\mathbf{C}$. Consequently, the constraint can be written as $\mathbf{C}_0^T(\mathbf{h} - \mathbf{b}) = 0$, and our objective function can be written as

$$\eta(\mathbf{a}, \lambda) = \mathbf{h}^T \mathbf{R} \mathbf{h} + \lambda^T \mathbf{C}_0^T (\mathbf{h} - \mathbf{b}), \tag{5}$$

where λ is the Lagrange multiplier vector. The optimum is at $\partial\eta / \partial\mathbf{h} = 0$ which yields $\mathbf{R}\mathbf{h} = \mathbf{C}_0 \lambda$. Parameter λ can be solved from $\mathbf{C}_0^T \mathbf{R}^{-1} \mathbf{C}_0 \lambda = \mathbf{C}_0^T \mathbf{b}$ [11].

Vector $\mathbf{a}$ obeys, by definition, relation $\mathbf{h} - \mathbf{b} = \mathbf{C}\mathbf{a}$ and can thus be calculated from $\mathbf{h}$. $\square$

Remark Since $e_n = \mathbf{x}^T \mathbf{h}$, vector $\mathbf{h}$ is the impulse response used in prediction. However, the minimal representation, containing no redundant data, is vector $\mathbf{a}$. It can be solved setting $E[\frac{\partial}{\partial \mathbf{a}} e_n^2] = 0$ from $\mathbf{C}\mathbf{R}\mathbf{C}^T \mathbf{a} = \mathbf{C}\mathbf{R}\mathbf{b}$.

The constrained linear predictors have m degrees of freedom in defining a predictor $\mathbf{h}$ of order $m+l-1$. This implies that in constrained predictors (for $l \geq 1$), in contrast to conventional LP, all-pole models of order $m+l-1$ can be defined from less than $m+l-1$ parameters [4, 5]. Consequently, constrained linear predictors can find applications in such areas as coding and feature extraction. However, covering all cases $l \geq 1$ is beyond the scope of this article and therefore we will concentrate on the case $l = 1$ only.

4. STABILITY OF CONSTRAINED LP MODELS

Theorem 1 *Predictor* $\mathbf{a}_c$ *solving*

$$\mathbf{R}\mathbf{a}_c = \gamma \left[1, c^{-1}, c^{-2}, \ldots, c^{-m}\right]^T \tag{6}$$

is minimum-phase if $|c| > 1$ *(where* $c \in \mathbb{C}$*),* $\mathbf{R}$ *is positive definite and* γ *is a scaling factor. For* $|c| = 1$ *predictor* $\mathbf{a}_c$ *has its roots on the unit circle, and for* $|c| < 1$ *outside the unit circle.*

The following proofs are based on the approach presented in [12].

Proof Let α be a root of polynomial $A_c(z)$ corresponding to vector $\mathbf{a}_c$. Writing vector $\mathbf{a}_c$ in factored form

$$\mathbf{a}_c = \begin{bmatrix} 1 \\ a_{c,1} \\ a_{c,2} \\ \vdots \\ a_{c,m} \end{bmatrix} = \begin{bmatrix} b_0 & 0 \\ b_1 & b_0 \\ b_2 & b_1 \\ \vdots & \vdots \\ 0 & b_{m-1} \end{bmatrix} \begin{bmatrix} 1 \\ -\alpha \end{bmatrix} = \mathbf{B} \begin{bmatrix} 1 \\ -\alpha \end{bmatrix}, \tag{7}$$

it follows that

$$\mathbf{R}\mathbf{a}_c = \mathbf{R}\mathbf{B} \begin{bmatrix} 1 \\ -\alpha \end{bmatrix} = \gamma \left[1, c^{-1}, c^{-2}, \ldots, c^{-m}\right]^T. \tag{8}$$

Multiplying from the left by $\mathbf{B}^H$ yields

$$\mathbf{B}^H\mathbf{R}\mathbf{B}\begin{bmatrix}1\\-\alpha\end{bmatrix} = \xi\begin{bmatrix}1\\c^{-1}\end{bmatrix}, \tag{9}$$

where $\xi = \gamma\sum_{i=0}^{m-1} b_i^* c^{-i}$.

If $\mathbf{R}$ is positive definite then $\mathbf{B}^H\mathbf{R}\mathbf{B}$ is positive definite since matrix $\mathbf{B}$ is full rank [13]. Further,

$$\mathbf{B}^H\mathbf{R}\mathbf{B} = \begin{bmatrix}s_0 & s_1\\ s_1^* & s_0\end{bmatrix} > 0, \tag{10}$$

which implies that $|s_0|^2 > |s_1|^2$ where s_0 is real.

Combining equations 9 and 10 yields

$$\mathbf{B}^H\mathbf{R}\mathbf{B}\begin{bmatrix}1\\-\alpha\end{bmatrix} = \begin{bmatrix}s_0-\alpha s_1\\ s_1^*-\alpha s_0\end{bmatrix} = \xi\begin{bmatrix}1\\c^{-1}\end{bmatrix} \tag{11}$$

Eliminating ξ to solve α yields

$$\alpha = \frac{s_1^* - s_0c^{-1}}{s_0 - s_1c^{-1}}. \tag{12}$$

We can readily show that $|\alpha| = 1$ is equivalent to

$$\left(|s_0|^2 - |s_1|^2\right)\left(1 - |c|^{-2}\right) = 0. \tag{13}$$

Therefore, $|\alpha| = 1$ requires that $|c| = 1$ since $|s_0| > |s_1|$. Further, in Eq. 11, $|c| > 1$ implies that $|\alpha| < 1$, while $|\alpha| \to \infty$ implies that $c \to \frac{s_0}{s_1}$, that is, $|c| < 1$. This concludes the proof. Note that we have, as a by-product, proved the stability of the traditional LP model. □

An example of the root space of predictor $\mathbf{a}_c$ as a function of c is depicted in Fig. 1.

Remark 1.1 Equation 6 can be rewritten as

$$\mathbf{R}\mathbf{a}_c = \gamma c^{-m}\left[c^m, c^{m-1}, \ldots, c^{-1}, 1\right]^T, \tag{14}$$

which shows that $\mathbf{a}_c^{\#}$ (where superscript $^{\#}$ denotes reversal of rows) is minimum-phase for $|c| < 1$. Thus, two constrained linear predictors defined by c and c^{-1} have their roots symmetric to the unit circle.

Remark 1.2 While Eq. 6 yields minimum-phase predictors and since an LP model with a simple FIR-filter constraint (Eq. 4) has a null-space $\mathbf{C}_0 = \left[1, c^{-1}, c^{-2}, \ldots, c^{-m}\right]^T$, the corresponding constrained LP model will be minimum-phase if and only if the zero of the FIR-filter is outside the unit circle.

Defintion The symmetric and antisymmetric part, $\mathbf{d}^+$ and $\mathbf{d}^-$, respectively, of a vector $\mathbf{d}$ are defined as $\mathbf{d}^\pm = \mathbf{d} \pm \mathbf{d}^\#$, where $^\#$ denotes reversal of rows. Similarly, the scaled (anti)symmetric part is $\hat{\mathbf{d}}^\pm = (\mathbf{d} \pm \mathbf{d}^\#)/(d_0 \pm d_m)$.

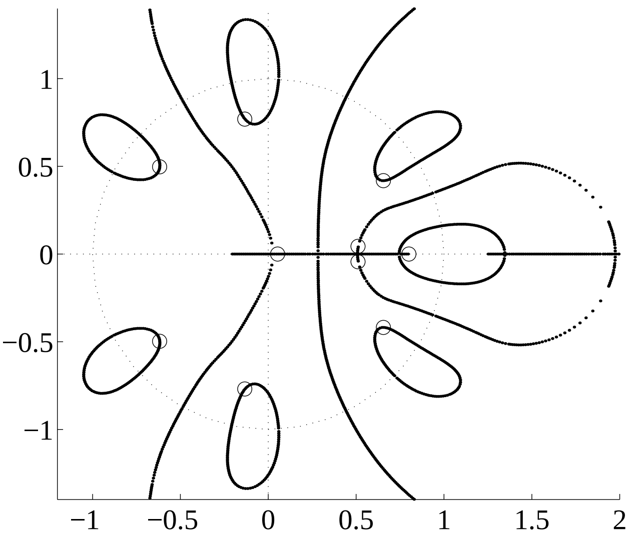

Fig. 1. Illustration of the root space of predictor $\mathbf{a}_c$ in Eq. 6 as a function of $c \in \mathbb{R}$ with $m = 10$. Corresponding LP-model ($c \to \infty$) is depicted with circles 'o'. When $|c| > 1$ roots of $\mathbf{a}_c$ lie within the unit circle, for $|c| = 1$ on the unit circle, and for $|c| < 1$ outside the unit circle.

Theorem 2 *The scaled symmetric and antisymmetric parts* $\hat{\mathbf{c}}^+$ *and* $\hat{\mathbf{c}}^-$*, respectively, of vector* $\mathbf{c} = \left[1, c^{-1}, \ldots, c^{-m}\right]^T$ *are, for* $c \to \pm 1$ *and* m *even*

$$\hat{\mathbf{c}}^+_{c\to+1} = [+1, +1, +1, \ldots, +1, +1]^T \tag{15}$$
$$\hat{\mathbf{c}}^+_{c\to-1} = [+1, -1, +1, \ldots, -1, +1]^T \tag{16}$$
$$\hat{\mathbf{c}}^-_{c\to+1} = [1, 1-\frac{2}{m}, 1-\frac{4}{m}, \ldots, \frac{2}{m}-1, -1]^T \tag{17}$$
$$\hat{\mathbf{c}}^-_{c\to-1} = [1, -1+\frac{2}{m}, 1-\frac{4}{m}, \ldots, -\frac{2}{m}+1, -1]^T. \tag{18}$$

Similarly, for m *odd*

$$\hat{\mathbf{c}}^+_{c\to+1} = [1, 1, 1, \ldots, 1, 1]^T \tag{19}$$
$$\hat{\mathbf{c}}^+_{c\to-1} = [1, -1+\frac{2}{m}, 1-\frac{4}{m}, \ldots, -\frac{2}{m}+1, 1]^T \tag{20}$$
$$\hat{\mathbf{c}}^-_{c\to+1} = [1, 1-\frac{2}{m}, 1-\frac{4}{m}, \ldots, \frac{2}{m}-1, -1]^T \tag{21}$$
$$\hat{\mathbf{c}}^-_{c\to-1} = [1, -1, 1, \ldots, 1, -1]^T. \tag{22}$$

The predictors $\mathbf{a}^\pm_{\pm c}$ *solving* $\mathbf{R}\mathbf{a}^\pm_{\pm c} = \gamma\hat{\mathbf{c}}^\pm_{c\to\pm1}$ *(where the sign of symmetry and* c *are not necessarily equal), have their zeros on the unit circle.*

Proof Equations 15, 16, 19 and 22, follow directly from the definition. Equations 17, 18, 20 and 21 can be readily derived using the l'Hopital rule for limits.

Since we have proved that predictors $\mathbf{a}_c$ solving Eq. 6 are minimum-phase if $|c| > 1$, then the symmetric and antisymmetric parts $\mathbf{a}_c^\pm$ have their zeros interlaced on the unit

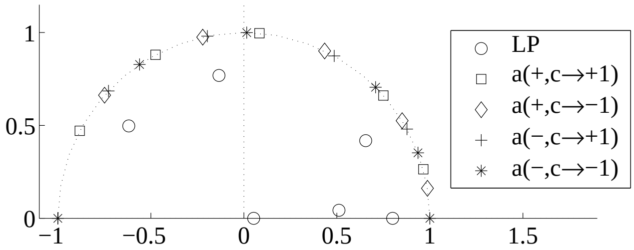

Fig. 2. An example of the unit circle and interlacing properties of the symmetric and antisymmetric parts $\mathbf{a}_c^{\pm}$ (Eqs. 15-22) of predictor $\mathbf{a}_c$ in Eq. 6 ($m = 10$).

circle [14]. The same holds for the reversed predictor $\mathbf{a}^{\#}$ by symmetry. When $|c| = 1$, the predictor $\mathbf{a}_c$ is either symmetric or antisymmetric and the opposite symmetry component is zero. However, it follows that at the limit $|c| \rightarrow 1$ the predictor $\mathbf{a}_{\pm c}^{\pm}$ must have its zeros on the unit circle since polynomial sequences are continuous. $\square$

The predictors $\mathbf{a}_c$ solving $\mathbf{R}\mathbf{a}_{\pm c}^{\pm} = \gamma\hat{\mathbf{c}}_{c\rightarrow\pm 1}^{\pm}$ are thus, in some combinations of sign of symmetry and sign of $c = \pm 1$, the LSP polynomials of degree $m+1$ with one trivial zero removed (see section 2.2). These polynomial have their zeros interlaced on the unit circle. In addition, symmetric and antisymmetric pairs with equal c are interlaced due to [14]. Further, we have observed that all these polynomials have their zeros interlaced on the unit circle. However, the proof of this statement lies beyond the scope of this article. The interlacing properties are demonstrated in Fig. 2. Notice especially how all combinations of predictor zeros are interlaced.

Due to the interlacing properties shown above, these theorems can be extended to multidimensional cases, i.e. cases where the FIR filter corresponding to Eq. 4 has several roots ($l > 1$), that is, when constraint vector in Eq. 6 is a Vandermonde matrix [13]. Unfortunately, again this is beyond the scope of this article.

5. CONCLUSIONS

This study deals with constrained linear prediction that differs from conventional LP in the sense that an optimal predictor of order $m + l - 1$ is defined using m parameters. In the present paper, we have focused solely on the stability properties of these models. These proofs form a basis for applications of constrained linear prediction which exist, for example, in such areas as speech coding and feature extraction.

REFERENCES

[1] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, no. 5, pp. 561–580, April 1975.

[2] G. Szegö, *Orthogonal Polynomials*, vol. 23, American Mathematical Society Colloquium Publications, New York, 3rd edition, 1967.

[3] Y. Bistritz, "Zero location with respect to the unit circle of discrete-time linear system polynomials," *Proc. IEEE*, vol. 72, no. 9, pp. 1131–1142, September 1984.

[4] S. Varho and P. Alku, "Separated linear prediction - a new all-pole modeling technique for speech analysis," *Speech Commun.*, vol. 24, pp. 111–121, 1998.

[5] P. Alku and S. Varho, "A new linear predictive method for compression of speech signals," in *Proc. International Conference on Spoken Language Processing (ICSLP)*, Sydney, Australia, 1998, vol. 6, pp. 2563–2566.

[6] W. B. Kleijn, T. Bäckström, and P. Alku, "On line spectral frequencies," *accepted for publication in IEEE Sig. Proc. Letters*, 2002.

[7] P. Alku and T. Bäckström, "All-pole modeling technique based on the weighted sum of the LSP polynomials," in *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'02*, Orlando, Florida, USA, May 2002, vol. 1, pp. 665–668.

[8] F. K. Soong and B.-H. Juang, "Line spectrum pair (LSP) and speech data compression," in *Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'84*, San Diego, CA, March 1984, vol. 1, pp. 1.10.1–1.10.4.

[9] G. A. Mian and G. Riccardi, "A localization property of line spectrum frequencies," *IEEE Trans. Speech Audio Proc.*, vol. 2, no. 4, pp. 536–539, 1994.

[10] J. Rothweiler, "On polynomial reduction in the computation of LSP frequencies," *IEEE Trans. Speech Audio Proc.*, vol. 7, no. 5, pp. 592–594, 1999.

[11] T. Söderstrom and P. Stoica, *System Identification*, Prentice Hall, New York, 1989.

[12] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*, J. Wiley & Sons, Inc., New York, 1996.

[13] G. Golub and C. van Loan, *Matrix Computations*, The Johns Hopkins University Press, London, 1996.

[14] H. W. Schüssler, "A stability theorem for discrete systems," *IEEE Trans. Acoust. Speech Signal Proc.*, vol. ASSP-24, no. 1, pp. 87–89, Feb. 1976.