MVDR BASED ALL-POLE MODELS FOR SPECTRAL CODING OF SPEECH

Manohar N. Murthi and Bhaskar D. Rao

Department of Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA 92039-0407 USA
manoharn@ece.ucsd.edu brao@ece.ucsd.edu

ABSTRACT

We present several analytical properties of Minimum Variance Distortionless Response (MVDR) based all-pole models that demonstrate the advantages and usefulness of these models for speech spectral coding. In particular, we show that a sufficient order MVDR all-pole model provides a spectral envelope that fits a set of spectral samples exactly with a parameterization convenient for quantization purposes. In addition, we show that MVDR all-pole filters provide a monotonically decreasing spectral distortion with increasing filter order. Furthermore, we show that the MVDR all-pole filter possesses the flexibility to be obtained from correlations based upon either spectral samples or conventional time-domain correlations. Finally, exploiting the insight gained from MVDR modeling, we introduce a novel class of constrained all-pole models for efficient spectral coding. In this approach, a subset of the Line Spectral Frequency (LSF) parameters associated with the all-pole model are judiciously fixed, leading to a simpler model parameterization.

1. INTRODUCTION

All-pole models are a popular parametric approach for modeling the short-term spectrum of speech, and for capturing the spectral envelope [1, 2]. The ability of all-pole filters to model the short-term spectrum of speech has led to their becoming a fundamental part of both time and frequency domain speech compression systems [3]. Linear Prediction (LP) is a popular method for obtaining the parameters of all-pole models with a long history and a rich knowledge base from which one can draw [4, 5]. However, it is well-known that Linear Prediction has its limitations in modeling voiced speech, particularly medium-pitched and high-pitched voiced speech. In particular, the LP spectrum tends to overestimate the spectral powers at the large valued pitch harmonics, providing a sharper contour than the original vocal tract response. In addition, increasing the LP model order does not lead to better spectral envelopes and often exacerbates the problem.

In [6, 7], we presented preliminary studies of Minimum Variance Distortionless Response (MVDR) spectrum modeling of voiced speech, and demonstrated the high order MVDR spectrum's superiority over LP spectral modeling. In particular, the high order MVDR spectrum was shown to perfectly model a voiced speech line spectrum. In this paper, we present some additional attractive analytical properties of MVDR spectral modeling of speech, and discuss the important issue of model parameterization for spectral coding. In particular, it is shown that MVDR models produce monotonically decreasing spectral distortions with increasing filter order. In addition, the MVDR spectrum that provides an envelope that fits a set of L voiced speech spectral powers exactly can be conveniently parameterized by $(L+1)$ parameters. Furthermore, we demonstrate that the MVDR spectrum can be computed from conventional time domain correlations. It can be shown that MVDR all-pole spectrum possesses the flexibility to model all types of speech including unvoiced speech [8]. Finally, exploiting the insight gained from MVDR modeling, we introduce a novel class of constrained all-pole models for efficient spectral coding. In this approach, a subset of the Line Spectral Frequency (LSF) parameters associated with the all-pole model are judiciously fixed, leading to a simpler model parameterization.

2. PROPERTIES OF MVDR ALL-POLE MODELING

The Mth order MVDR all-pole spectrum is also known as the Capon spectrum [9, 10] and is given by

$$P_{MV}^{(M)}(\omega) = \frac{1}{\mathbf{v^H}(\omega)\mathbf{R_{M+1}^{-1}}\mathbf{v}(\omega)} = \frac{1}{|B(e^{j\omega})|^2}, \qquad (1)$$

where $\mathbf{v}(\omega) = [1, e^{j\omega}, e^{j2\omega}, \cdots, e^{jM\omega}]^T$, and $\mathbf{R_{M+1}}$ is the $(M+1)\times(M+1)$ Toeplitz autocorrelation matrix of the input speech, and $1/B(z)$ is the associated MVDR all-pole filter. The MVDR spectrum can be obtained by a simple non-iterative computation involving the Linear Prediction Coefficients while the associated stable and causal MVDR all-pole filter $1/B(z)$ can be obtained by a spectral factorization [8].

For modeling voiced speech, the MVDR all-pole spectrum has several properties that make it an attractive candidate for spectral coding of speech. To illustrate these analytical properties, we consider a voiced speech signal modeled as a sum of harmonic signals

$$u(n) = \sum_{k=1}^{L} c_k \ \cos(\omega_0 kn + \phi_k), \qquad (2)$$

where $\omega_0 = 2\pi f_0$ with f_0 the pitch frequency of the speech, where c_k's are the amplitudes at the harmonics, and where L is the number of harmonics given by $L = \lfloor \frac{f_s}{2f_0} \rfloor$ with f_s the sampling frequency, typically 8kHz. Such a voiced speech signal has L harmonics or alternately $2L$ exponentials at the positive and negative multiples of the fundamental pitch frequency. With this model for

This work was supported by UC MICRO Grants 97-146, 98-125, Hughes Electronics, Nokia, and Qualcomm

voiced speech, the signal has a correlation sequence

$$r_{uu}(m) = \sum_{k=1}^{L} \frac{|c_k|^2}{2} \cos(\omega_0 k m). \quad (3)$$

The power spectrum exhibits a discrete line spectrum with powers $S_k = |c_k|^2/4$ at the harmonic frequencies $\omega_0 k$.

With this harmonic model of voiced speech, we can make the following observation.

Theorem 1 *For filter orders $M < (2L-1)$, the MVDR spectrum satisfies $P_{MV}^{(M-1)}(\omega_0 l) \geq P_{MV}^{(M)}(\omega_0 l) \geq S_l$. The sequence $(P_{MV}^{(M)}(\omega_0 l) - S_l)$ decreases monotonically as M increases, reaching zero for $M \geq (2L-1)$. For filter orders of $M \geq (2L-1)$, $P_{MV}(\omega_0 l) = S_l$.*

Proof: See [6] for the basic ideas. See [8] for a more thorough discussion. □

Therefore the MVDR all-pole spectrum with sufficient filter order can model a set of voiced speech spectral samples exactly. Furthermore, MVDR all-pole spectral modeling of voiced speech improves as the filter order increases. The improved modeling of Theorem 1 can be stated in terms of local spectral distortion measures. We consider local spectral distortion measures $\mathbf{D}(x, \hat{x})$

which measure distortion between a true spectral value x and an estimate $\hat{x}$ at a particular frequency, such as $[\log(\frac{x}{\hat{x}})]^2$.

Theorem 2 *The local harmonic spectral distortion*[1] *at $\omega_0 l$, $\mathbf{D}(P_{MV}^{(M)}(\omega_0 l), S_l)$, decreases monotonically as the filter order M increases, i.e., $\mathbf{D}(P_{MV}^{(M)}(\omega_0 l), S_l)) \leq \mathbf{D}(P_{MV}^{(M-1)}(\omega_0 l), S_l)$. The spectral distortion is zero for $M \geq (2L-1)$, or $\mathbf{D}(P_{MV}^{(2L-1)}(\omega_0 l), S_l) = 0$.*

Proof: See [8] □

It follows that any global spectral distortion which is a linear combination of local harmonic distortions must also monotonically decrease. Having established this superior ability of high order MVDR models to model voiced speech, we now consider the parameterization of these models. We show that the MVDR spectrum of order $(2L-1)$ that models L harmonics can still be parameterized by L parameters.

Theorem 3 *The MVDR spectrum of order $M = (2L-1)$ that is used to model perfectly a voiced speech discrete line spectrum of L harmonics can be completely parameterized by (L+1) parameters: (i) a fundamental pitch value, (ii) $(L-1)$ Line Spectral Frequencies (LSFs), and (iii) a gain value.*

Proof: The MVDR spectrum of order $M = (2L-1)$ is completely specified by the LP filter $A_M(z)$ of order $M = (2L-1)$ and the prediction error variance P_e [10]. The LSFs of $A_M(z)$ are given by the roots of the polynomials of $P(z) = A_M(z) + z^{-2L}A_M(z^{-1})$ and $Q(z) = A_M(z) - z^{-2L}A_M(z^{-1})$. Because $P(z)$ and $Q(z)$ have real coefficients, only the roots between $\omega = 0$ and π need to be considered. From linear prediction theory (see [8]), $P(z)$, which results from one step of the Levinson-Durbin recursion, has its complex conjugate roots on the unit circle with L roots at the pitch harmonic frequencies between 0 and π. Therefore only the pitch frequency is needed to specify $P(z)$. $Q(z)$ has roots at $\omega = 0$ and π with $(L-1)$ roots between $\omega = 0$ and π. Only the $(L-1)$ roots between 0 and π are needed to specify $Q(z)$. In total, the pitch frequency, the $(L-1)$ roots of $Q(z)$ between 0 and π, and the prediction error variance, P_e, are needed to completely specify the MVDR spectrum of order $M = (2L-1)$. □

Therefore, the MVDR all-pole spectrum of order $M = (2L-1)$ that models L voiced speech harmonics exactly does not require a large amount of parameterization for perfect modeling. Furthermore, most of the required parameters are LSFs which are already used in all-pole model parameterization and quantization.

Note that many of the properties of the MVDR spectrum do not depend on the harmonic structure of the input signal and in fact the results in this section can be generalized to line spectra with arbitrary frequency patterns. In fact, the MVDR spectrum of order $M = (2L-1)$ models a symmetric discrete line spectrum consisting of L spectral samples at L distinct frequencies exactly [8]. If the spectral sampling scheme is known, then a low parameterization similar to the one in Theorem 3 can be obtained [8].

[1]$\mathbf{D}$ includes most reasonable distortion measures. For a more precise characterization of allowable $\mathbf{D}$, see [8]

3. PERFORMANCE OF MVDR MODELING OF SPEECH

Now we show some examples illustrating the properties of MVDR all-pole models. First, we consider the case of modeling a set of spectral samples, an important task in spectral coding of speech.

3.1. MVDR modeling of speech spectral samples

In these examples, the MVDR spectrum is computed based on correlations in Eq. 3 in which the spectral samples are determined from conventional peak-picking methods.

Example 1:*Perfect MVDR modeling of non-uniformly spaced spectral samples*. In this example, 10 non-uniformly spaced spectral samples are used to form a discrete line spectrum at both positive and negative frequencies. The 10 samples are modeled by both MVDR and LP spectra (with real filter coefficients) of filter order $M = 19$. The results of this experiment are illustrated in Figure 1. The MVDR spectrum provides an all-pole envelope that fits the 10 spectral samples exactly while the LP spectrum does not provide good estimates of the spectral samples.

Example 2: *MVDR Spectral Sample Modeling for Suboptimal M* For this example, a fixed voiced speech spectrum was sampled for a pitch frequency of 280 Hz, resulting in $L = 14$ harmonics. The MVDR all-pole spectral estimates of filter orders $M = 10$ to $(2L-1) = 27$ are compared to the true spectral powers by a Discrete Log-Spectral Distortion $LSD = [\frac{1}{L}\sum_{m=1}^{L} [10\log_{10} P(\omega_0 m) - 10\log_{10} \hat{P}(\omega_0 m)]^2]^{1/2}$ where L is the number of harmonics, $P(\omega)$ corresponds to the true spectral powers, and $\hat{P}(\omega)$ corresponds to the estimated spectral powers. The results are illustrated in Figure 2. As expected, the MVDR discrete spectral distortion monotonically decreases as the filter order increases, with perfect modeling when the filter order $M = (2L-1) = 27$. Such a filter can be characterized by $(L+1) = 15$ parameters.

Furthermore, the MVDR estimates for the three harmonics with the largest powers in the voiced speech line spectrum were compared to the true powers, and local Log Spectral Distortions were computed. The results are in Figure 3. The MVDR spectral estimates have an initial positive bias that decays to zero as the filter order increases, resulting in monotonically decreasing spectral distortion. The LP spectral estimate distortions are also shown, and tend to get very large as the filter order is increased, resulting in worsening spectral distortion.

3.2. MVDR spectrum based on time domain correlations

In the previous section, the MVDR all-pole spectrum was used to provide a perfect all-pole envelope fit to set of spectral samples. In this section, we briefly consider MVDR all-pole models that are based on time-domain correlations. More extensive discussion and performance analysis can be found in [8]. In these examples, biased correlation estimates of the speech, $\hat{r}(0), \cdots, \hat{r}(M)$ are computed for a block of speech of $N = 160$ samples, i.e. $\hat{r}_{xx}(m) = \frac{1}{N}\sum_{n=0}^{N-m-1} x(n+m)x(n), 0 \le m \le N-1$. The MVDR spectrum is computed from the Linear Prediction Coefficients based upon these time-domain correlations.

Example 3:*MVDR Modeling of 4 Voiced Spectra with Different Pitch Frequencies*. In this experiment, we compute the MVDR spectrum for four different voiced speech spectra of varying pitch frequencies. The results are shown in Figure 4. In all cases, the MVDR spectrum has an order roughly two times the number of pitch harmonics. The MVDR spectrum offers a smooth contour that provides a good estimate of the spectral powers at the harmonics.

The MVDR spectrum is also applicable to unvoiced speech and mixed spectra and need not be based upon frequency domain spectral sampling methods. For a more comprehensive performance study of MVDR spectral modeling of speech, see [8].

4. ALL-POLE MODELS WITH CONSTRAINED LSF LOCATIONS

For low bit rate speech coders, primarily harmonic coders, quantizing $(L-1)$ LSFs, a pitch value, and a gain that characterize the MVDR all-pole envelope that perfectly models L spectral samples is suitable for high pitch speech. For medium and low pitch speech, all-pole models with simpler parameterization are desirable.

In this section we introduce a novel class of constrained all-pole models. The constraints in these models are in terms of the location of a subset of the LSFs associated with the model. By deliberately constraining the locations of LSFs, the number of parameters required for the characterization of the all-pole model is reduced. For a detailed study of all-pole models with constrained LSFs, see [11]. The particular class of all-pole filters introduced here is based upon the MVDR spectrum and exploits the property highlighted in Theorem 3. In particular, if we are allowed a model with only $(P+1)$ parameters, $P < L$, then from Theorem 3 the MVDR spectrum of order $M = 2P-1$ parameterized by $(P+1)$ parameters can model only P harmonic exactly. We develop an approach that uses this property and tries to model all the harmonics in an optimal manner. In this constrained LSF MVDR approach, the frequency of the P harmonics are fixed and their power levels determined to ensure that the resulting MVDR spectrum sampled at the actual L harmonic locations matched the power of the L harmonics in an optimal manner. Fixing the location of the P harmonics in a predetermined manner constrains the locations of the LSF (c.f. Theorem 3). We now elaborate on this approach.

The P spectral powers $|a_1|^2/4, \cdots, |a_P|^2/4$ at P harmonic frequencies $\omega_c, 2\omega_c, \cdots, P\omega_c$ where $\omega_c = \pi/(P+1)$ determine the constrained LSF MVDR spectrum $P_{CL-MV}(\omega)$ of order $(2P-1)$. The P harmonics have a correlation sequence of $r_P(m) = \sum_{k=1}^{P} \frac{|a_k|^2}{4}\cos(\omega_c k m)$, and the Constrained LSF MVDR spectrum of order $M = 2P-1$ is specified by

$$P_{CL-MV}(\omega) = \frac{1}{\mathbf{v}^H(\omega)\mathbf{R}_{(\mathbf{2P})}^{-1}\mathbf{v}(\omega)} \tag{4}$$

where $\mathbf{R}_{(\mathbf{2P})}$ is the $2P \times 2P$ Toeplitz Autocorrelation matrix formed with correlation sequence $r_P(m)$.

The P spectral powers are selected such that $P_{CL-MV}(\omega)$ sampled at the L voiced speech harmonic frequencies is approximately equal to the L true voiced speech spectral powers, or

$$P_{CL-MV}(\omega) \approx \frac{|c_l|^2}{4}, l = 1 \cdots L. \tag{5}$$

In order to compute the P spectral powers, we need to carry out some manipulations. The matrix $\mathbf{R}_{(\mathbf{2P})}$ can be decomposed as $\mathbf{R}_{(\mathbf{2P})} = \mathbf{W_c}\mathbf{D_A}\mathbf{W_c}^H$ where $\mathbf{W_c}$ is a $2P \times 2P$ matrix with $\mathbf{W_c} = [\mathbf{V_c}|\mathbf{V_c^*}]$ and where $\mathbf{V_c} = [\mathbf{v}(\omega_c)\mathbf{v}(2\omega_c)\cdots\mathbf{v}(P\omega_c)]$, in which $\mathbf{v}(\omega)$ is defined as in Eq. 1, and $\mathbf{D_A}$ is a $2P \times 2P$ diagonal matrix with $\mathbf{D_A} = [\begin{smallmatrix}\mathbf{\Lambda_A} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda_A}\end{smallmatrix}]$ where $\mathbf{\Lambda_A}$ is a diagonal matrix of P unknown spectral parameters with $(\mathbf{\Lambda_A})_{ii} = |a_i|^2/4$.

Utilizing the matrix decomposition, the Constrained LSF MVDR spectrum's desired property in Eq. 5 has the form

$$\mathbf{v}^H(l\omega_0)\mathbf{W_c^{-H}}\mathbf{D_A^{-1}}\mathbf{W_c^{-1}}\mathbf{v}(l\omega_0) \approx \frac{4}{|c_l|^2}, l = 1 \cdots L. \tag{6}$$

If we define $\mathbf{u}(l\omega_0;\omega_c) = \mathbf{W_c}^{-1}\mathbf{v}(l\omega_0)$, we can rewrite the desired property as

$$\mathbf{u}^H(l\omega_0;\omega_c)\mathbf{D_A}^{-1}\mathbf{u}(l\omega_0;\omega_c) \approx \frac{4}{|c_l|^2}, l = 1 \cdots L.. \tag{7}$$

Putting together the desired properties for all L harmonics into matrix form results in

$$\mathbf{Qa} \approx \mathbf{c}, \tag{8}$$

where $\mathbf{Q}$ is an $L \times P$ matrix with elements $\mathbf{Q_{ij}} = (|u_j(i\omega_0;\omega_c)|^2 + |u_{(j+P)}(i\omega_0;\omega_c)|^2)$ in which $u_j(i\omega_0;\omega_c)$ is the jth element of the vector $\mathbf{u}(i\omega_0;\omega_c) = \mathbf{W_c}^{-1}\mathbf{v}(i\omega_0)$ defined above, $\mathbf{a}$ is a $P \times 1$ vector of the reciprocals of the unknown spectral powers, i.e. $\mathbf{a_i} = 4/|a_i|^2, i = 1, \cdots, P$, and $\mathbf{c_i} = 4/|c_i|^2, i = 1, \cdots, L$ is a $L \times 1$ vector of the reciprocals of the true voiced speech spectral powers.

For this overdetermined system of L equations in P unknowns, incorporating perceptual considerations, $\mathbf{a}$ can be obtained as a solution to the weighted least squares problem [11]

$$\min_{\mathbf{a}} \|\mathbf{U}(\mathbf{Qa} - \mathbf{c})\|_2, \tag{9}$$

where U is a $L \times L$ diagonal weighting matrix with the L voiced speech spectral powers along the diagonal, i.e. $\mathbf{U_{ii}} = |c_i|^2/4$.

Example 4:*Example of Constrained LSF MVDR Modeling*. We consider the case of a voiced speech signal consisting of $L = 24$ harmonics. The Constrained LSF MVDR model is characterized by $P = 18 + 1 = 19$ parameters. The results are shown in Figure 5. The Constrained LSF MVDR does a reasonable job of modeling the voiced speech harmonics.

For many cases, preliminary results indicate that Constrained LSF MVDR spectra with parameterization set to $0.8L < P < L$ provide reasonable modeling. A more detailed study of Constrained LSF MVDR and other model order reduction techniques can be found in [11].

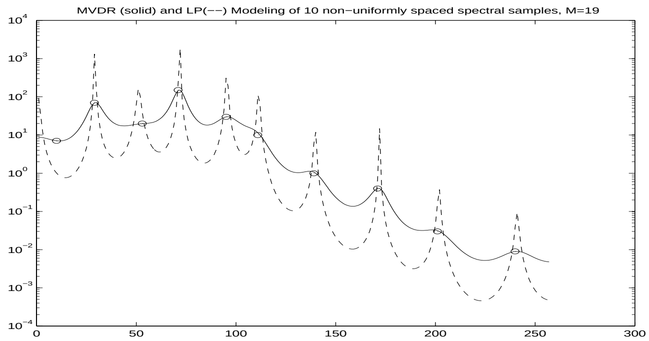

Figure 1: MVDR (solid) and LP (dashed) Spectral Modeling of 10 non-uniformly spaced spectral samples with Filter Orders of M=19. From *Example 1.*

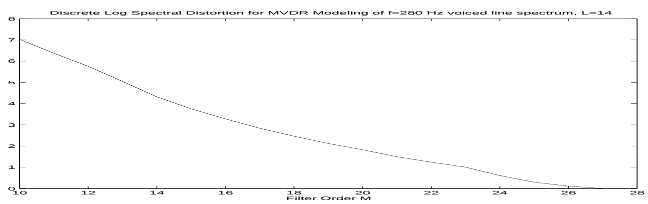

Figure 2: Comparison of MVDR Spectral Distortions for various filter orders in modeling of a synthetic 280 Hz voiced speech spectrum with $L = 14$ harmonics. From *Example 2.*

5. REFERENCES

[1] L.R. Rabiner and R.W. Schafer. *Digital Processing of Speech Signals*. Prentice-Hall, Englewood Cliffs, NJ, 1978.

[2] D. O'Shaughnessy. *Speech Communication: Human and Machine*. Addison-Wesley, 1987.

[3] W. B. Kleijn and K. K. Paliwal. *Editors, Speech Coding and Synthesis*. Elsevier, 1995.

[4] J.D. Markel and A.H. Gray Jr. *Linear Prediction of Speech*. Springer-Verlag, New York, 1976.

[5] J. Makhoul. "Linear prediction - A tutorial review". *Proceedings IEEE*, 63(4):501–580, April 1975.

[6] M. N. Murthi and B. D. Rao. "Minimum Variance Distortionless Response (MVDR) Modeling of Voiced Speech". In *Proc. ICASSP 97*, Munich, Germany, April 1997.

[7] M. N. Murthi and B. D. Rao. "All-Pole Model Parameter Estimation for Voiced Speech". In *Proc. 1997 IEEE Workshop on Speech Coding For Telecomm.*, 1997.

[8] M. N. Murthi and B. D. Rao. "All-Pole Modeling of Speech Based on the Minimum Variance Distortionless Response Spectrum". *Center for Wireless Communications Report No. 9803, University of California, San Diego.*

[9] J. Capon "High-resolution frequency-wavenumber spectrum analysis" *Proceedings of the IEEE*, vol. 57, (no. 8):1408-18, August 1969.

[10] S. Haykin. *Adaptive Filter Theory*. Prentice-Hall, Englewood Cliffs,NJ, 1991.

[11] M. N. Murthi. Ph.D. thesis, in preparation. *Department of Electrical and Computer Engineering, University of California, San Diego.*

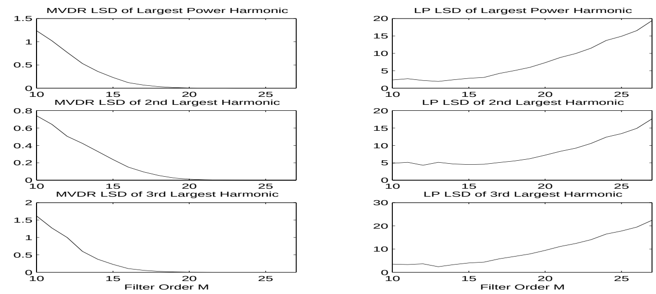

Figure 3: Comparison of MVDR and LP Spectral Distortions in modeling the 3 largest power voiced harmonics. From *Example 2.*

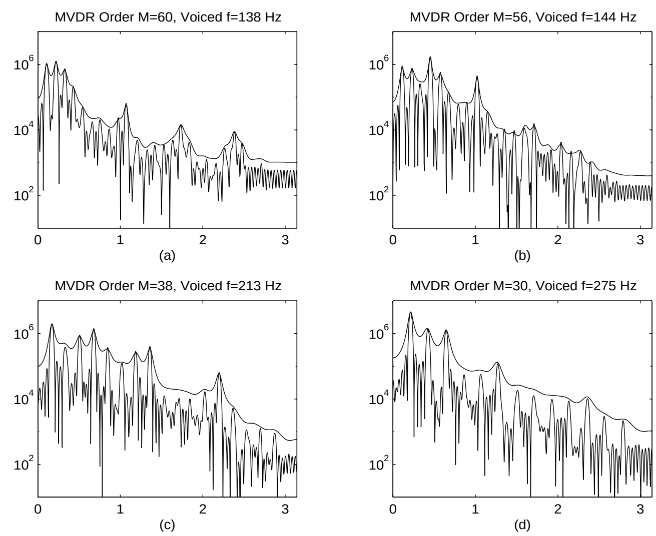

Figure 4: MVDR spectral modeling of four different voiced speech spectra. The MVDR spectra are computed from conventional time-domain correlations of frames of $N = 160$ samples of speech (a) M=60, f=138Hz (b) M=56, f=144Hz (c) M=38, f=213 Hz (d) M=30 f=275. From *Example 3.*

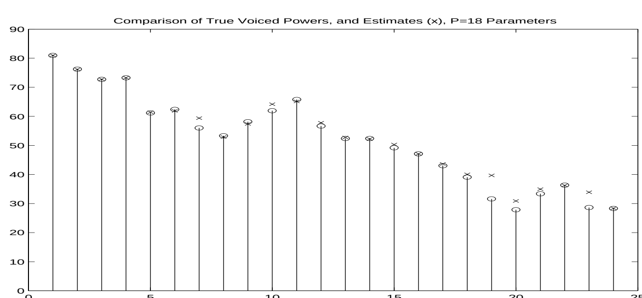

Figure 5: Comparison of True Spectral Powers (O), and Constrained LSF MVDR power estimates (x), with number of quantization parameters P=18. From *Example 4.*