POLE ZERO ESTIMATION FROM SPEECH SIGNALS BY AN ITERATIVE PROCEDURE

K. Schnell, A. Lacroix

Institut für Angewandte Physik, J.W. Goethe-Universität
Robert-Mayer Str. 2-4
D-60325 Frankfurt am Main, Germany
email: Lacroix@iap.uni-frankfurt.de

ABSTRACT

An iterative procedure is discussed to estimate poles and zeros of a rational transfer function from speech signals, which takes advantage of the individual solutions of AR and MA processes. Besides speech, analyses of test signals are also carried out, which lead to optimal results. In contrast to Prony's and related methods, the algorithm don't presuppose a pair of an input and output signal. The proposed procedure is specialised for the analysis of periodic signals, though it can be applied to non-periodic signals, too. The algorithm combines two known partial solutions in an iterative way. The estimation of an all-pole model is obtained by the Burg method and the estimation of zeros by using the inverted signal in the spectral domain. It can be shown that the power spectrum of analyzed speech periods can be better modelled by poles and zeros especially with respect to the gaps in the spectrum.

1. INTRODUCTION

For all-pole models the linear prediction, realised by the Burg method or the Levinson-Durbin recursion, provides a robust and good estimation [1]. For pole-zero models, there are no comparable algorithms. Algorithms for pole zero estimation exist with Prony's and related methods, which assume an input and output signal of the analyzed model. A special case of this is the blind deconvolution. If the input is unknown and only the output is available, an estimated impulse response $h'(n)$ of the analyzed output signal $x(n)$ can be modeled by the impulse response $h(n)$ of an LTI system having poles and zeros of the rational transfer function $H(z) = B(z)/A(z)$ [2,3]. Therefor the problem is to minimize the following expression:

$$\sum_{k=0}^{K} \left(h'_k - h_k\right)^2 \quad \rightarrow \quad \min \ . \qquad (1)$$

Shank's method, Kalman's method and iterative prefiltering proposed by Steiglitz treat the minimization (1) in different ways [4]. The impulse response should be estimated at first, if Prony's and related methods are used. An another approach is based on Durbin's method for MA processes, which estimates the zeros by the solution of an high-order all-pole model. This is used in Durbin's second method for the estimation of poles and zeros. The implementation of Durbin's second method can vary [5,6]. In this contribution the proposed procedure is based on inverse filtering. It combines two solutions of an AR process and a MA process, which represent partial solutions in this algorithm. A first attempt is started on already in [7], by an alternate use of these two solutions, and in [8] using the power density spectrum. In the following an improved iterative procedure is described.

2. ITERATIVE PROCEDURE

2.1 Error Definition

For the estimation of the poles and zeros a recursive prediction filter is used. $\hat{x}(n)$ is the prediction of the analyzed signal $x(n)$

$$e(n) = x(n) - \hat{x}(n) = x(n) - \sum_{k=1}^{N} a_k \cdot x(n-k) + \sum_{k=1}^{M} b_k \cdot e(n-k) \, .$$

The problem is to minimize the power of the prediction error $e(n)$. This can be formulated in the frequency domain:

$$\varepsilon = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|E(\mathrm{e}^{\mathrm{j}\omega})\right|^2 \mathrm{d}\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|\frac{1-\sum_{k=1}^{N} a_k \mathrm{e}^{-\mathrm{j}\omega k}}{1-\sum_{k=1}^{M} b_k \mathrm{e}^{-\mathrm{j}\omega k}} \cdot X(\mathrm{e}^{\mathrm{j}\omega})\right|^2 \mathrm{d}\omega \quad .(2)$$

$A(z)$ is the denominator of the transfer function $H(z)$ and represents the poles of the model, respectively $B(z)$ is the numerator and represents the zeros:

$$H(z) = \frac{1-\sum_{k=1}^{M} b_k z^{-k}}{1-\sum_{k=1}^{N} a_k z^{-k}} = \frac{B(z)}{A(z)} \quad .$$

Generally $\varepsilon(A,B)$ cannot be minimized directly. However, if the zeros or poles are known the remaining coefficients can be obtained straightforward. These two partial solutions are described in the following section.

1.2 Partial Solution I

If the zeros are known, the optimal poles can be determined by applying the Burg method to the analyzed signal, in which the known zeros are removed. If the zeros are not removed, the zeros distort the estimation process because the estimated poles tend to approximate not only the actual poles but also the zeros of the model. However, the optimal zeros usually are unknown. Therefore an approximation $\overline{B}(z)$ of the optimal numerator coefficients is used. The denominator coefficients a_k can be estimated with aid of $\overline{B}(z)$ by minimization of the error:

$$\varepsilon_1 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\left(1-\sum_{k=1}^{N} a_k \mathrm{e}^{-\mathrm{j}\omega k}\right)\cdot\frac{X(\mathrm{e}^{\mathrm{j}\omega})}{\bar{B}(\mathrm{e}^{\mathrm{j}\omega})}\right|^2 \mathrm{d}\omega \quad . \qquad (3)$$

Now the Burg method is applied to the IDFT of $X(z)/\bar{B}(z)$, which implies that the assumed zeros are reduced in the spectral domain and cannot anymore distort the estimation process. The better the zeros are approximated the estimation of the poles is improved.

2.2 Partial Solution II

For the partial solution II corresponding to the partial solution I the numerator $B(z)$ is estimated by using an assumed approximation $\bar{A}(z)$ of the denominator. Since the estimated coefficients are now in the denominator, the Burg method cannot be applied to this form of the error. But if the integrand in the error is inverted, the poles are changed into zeros. This error definition

$$\varepsilon_2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\left(1-\sum_{k=1}^{M} b_k \mathrm{e}^{-\mathrm{j}\omega k}\right)\frac{1}{\bar{A}(\mathrm{e}^{\mathrm{j}\omega})\cdot X(\mathrm{e}^{\mathrm{j}\omega})}\right|^2 \mathrm{d}\omega \qquad (4)$$

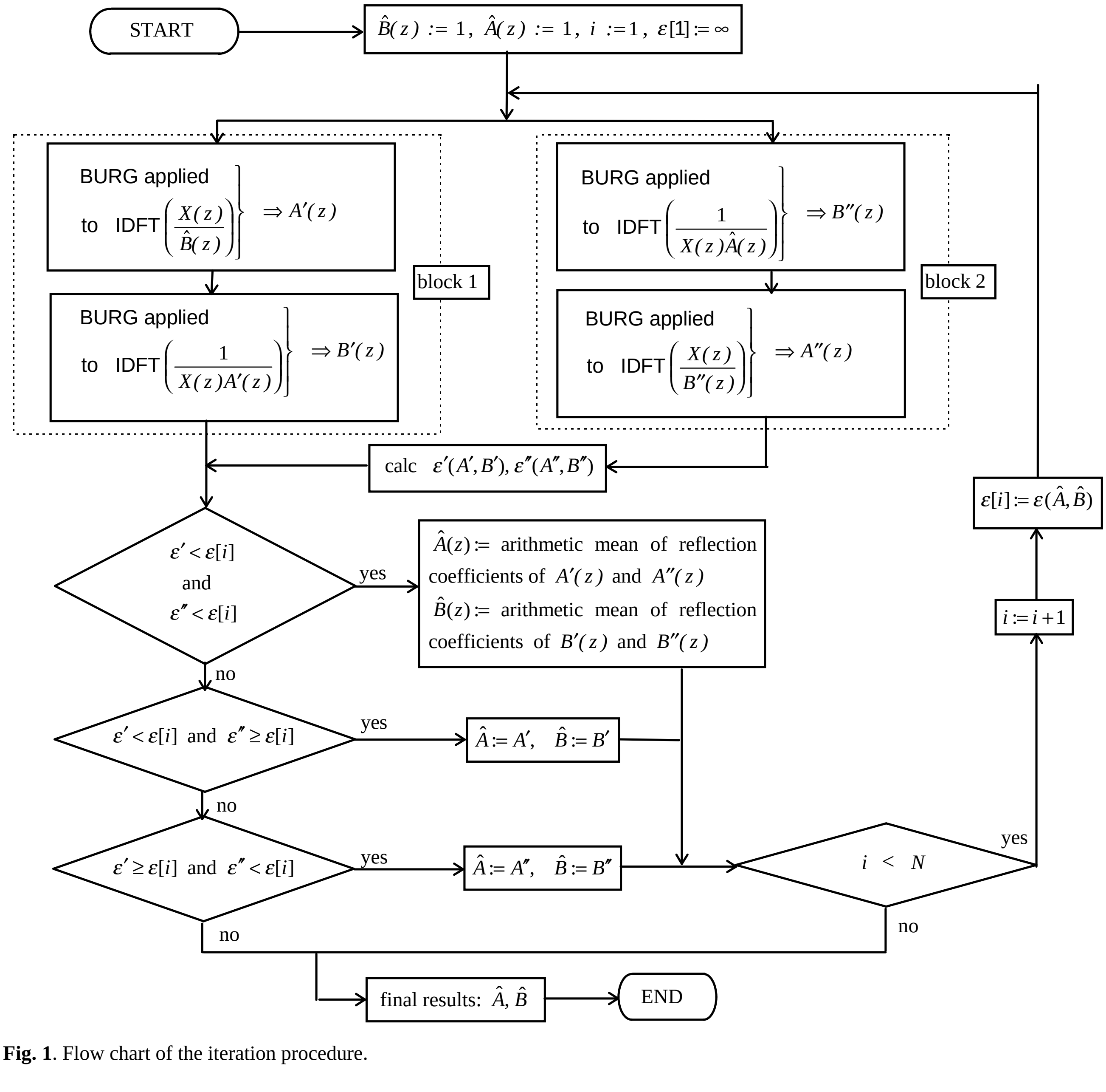

Fig. 1. Flow chart of the iteration procedure.

allows to estimate the coefficients b_k by the Burg method. To minimize ε_2 the Burg method is applied to the IDFT of $\overline{A}^{-1}(z) \cdot X^{-1}(z)$. The assumed poles are reduced in the spectral domain, corresponding to the partial solution I.

2.3 Complete Solution

These two partial solutions are not independent of each other, because one partial solution requires the results of the other partial solution. Therefore the solutions should be executed alternately. The results of the previous partial solutions are the assumptions of the next partial solutions. The succession of the partial solutions cannot be arranged any way, because the results are different whether the poles are estimated after the estimation of the zeros or vice versa. Therefore the procedure executes two blocks parallel, in which two partial solutions are carried out one after the other. In one block the poles are estimated after the zeros and in the another block the order is reverse. In block 1 the estimated poles $A'(z)$ are used for the following estimation of $B'(z)$ and corresponding in block 2 the estimated zeros $B''(z)$ are used for the following estimation of $A''(z)$. Afterwards the errors $\varepsilon'(A',B')$ with the results of block 1 and $\varepsilon''(A'',B'')$ with the results of block 2 are calculated. If both errors are larger than the error of the previous iteration the procedure breaks and the final solution is the result of the previous iteration. If one error of one block is larger and one is smaller than the error of the previous iteration, the resulting error of this iteration is the result of the block with the smaller error. If the two errors of the blocks are smaller than the error of the previous iteration, the solution of the iteration is a combination of the results of the two blocks. This is realized by the arithmetic mean of two sets of reflection coefficients, which are obtained from $A'(z)$ and $A''(z)$. The conversion of the resulting reflection coefficients into the polynomial coefficients represents the new estimated coefficients $\hat{a}_k$. Corresponding operations are carried out for the $\hat{b}_k$. Because of the use of the reflection coefficients, the stability of the resulting system is guaranteed. Eventually the iteration finishes, when a number N of iterations is exceeded. The iterative procedure begins with the start configuration $\hat{A}(z)=1$ and $\hat{B}(z)=1$, which implies no knowledge about the poles and zeros of the system. The flow chart of the whole algorithm is depicted in figure 1.

2.4 Burg Method used for Periodic Signals

For periodic signals a special modification of the Burg method has been developed. The calculation of the correlation function for the Burg coefficient requires values, which are outside of the analyzed segment. These values can be declared to zero. If a periodic signal is assumed, the outside value can be described with a value, which is inside of the analyzed segment, by considering the periodicity. Therefor it is favorable, that the segment is one period of the analyzed signal.

3. EXAMPLES

3.1 Examples from Test Signals

To demonstrate the capability of the procedure, test signals are generated, which are produced by prescribed systems. E.g. a system with 10 poles and 4 zeros is excited by an impulse train. A period of the output signal is analyzed. For the transfer function during the iteration the same number of poles and zeros is chosen, so that it is possible that the procedure can describe the analyzed model precisely. Figure 2 shows the result after the first iteration. The line spectrum is the DFT of the analyzed test

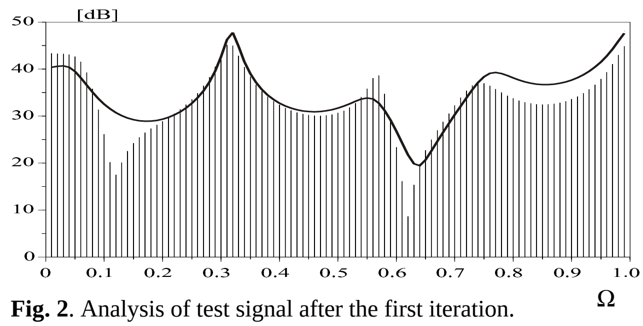

Fig. 2. Analysis of test signal after the first iteration.

signal and the solid line is the magnitude response of the estimated transfer function. After the first iteration the magnitude response is an inadequate approximation of the DFT spectrum of the test signal. The procedure finishes after 15 iterations. After that iteration the estimated magnitude response in figure 3 gives a perfect match to the DFT spectrum of the

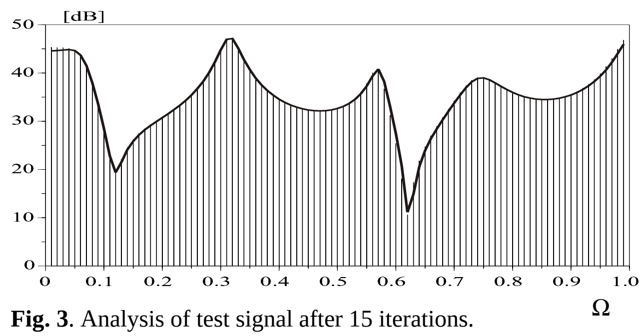

Fig. 3. Analysis of test signal after 15 iterations.

analyzed period. More test signals have been analyzed showing comparable results.

3.2 Examples from Speech Signals

In practice speech signals cannot be perfectly modeled in contrast to the test signals due to the imperfections of the model. However, the dominant resonances and existing antiresonances in the speech spectrum should be modelled. In figure 4 results of the analysis of one period of the nasal /n/ is shown. The order of

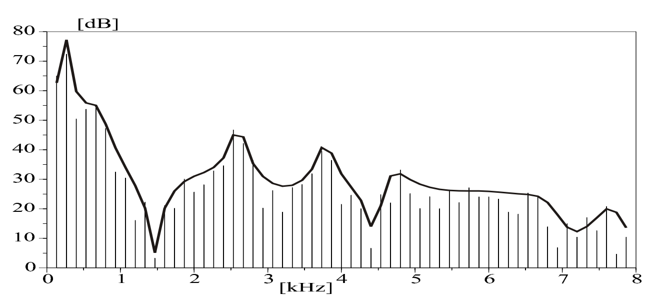

Fig. 4. Analyzed period of /n/ by 20 poles and 10 zeros.

the analyzed system is 20 for the poles in combination with 10 zeros. The line spectrum represents the DFT of the analyzed speech period and the solid lines are the estimated magnitude responses. The zero at 1500 Hz can be well observed in figure 4, which is caused by the coupling of the mouth cavity with the nasal tract. Due to the fluctuations of the glottal source and other effects single periods of vowels show also gaps in the magnitude spectrum. This fact can be seen in figure 5 showing the result of the analysis of one period of the vowel /i:/; the estimated system has 20 poles and 10 zeros. For comparison figure 6 shows the analysis of the same speech period /i:/ with 30 poles but no zeros. Although more poles have been spent, the gaps in the spectrum cannot be modelled by an all-pole system, because of the missing

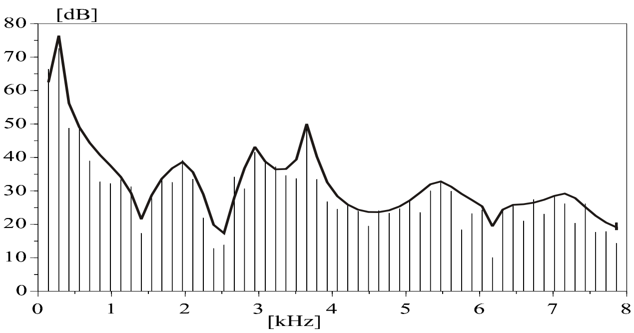

Fig. 5. Analyzed period of /i:/ by 20 poles and 10 zeros.

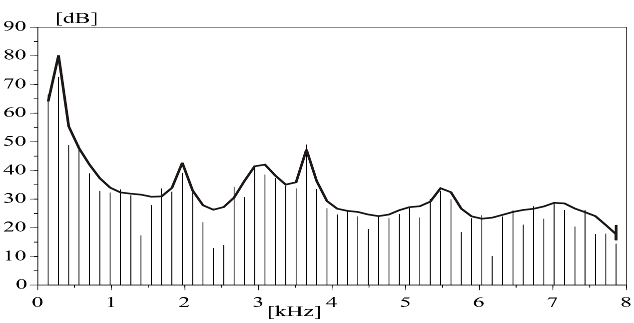

Fig. 6. Analyzed period of /i:/ by 30 poles but no zeros.

zeros. Figure 7 shows an example /z/ of a voiced fricative. The fluctuations of the adjacent periods of /z/ are especially strong due to the additional noisy excitation of this sound. For the

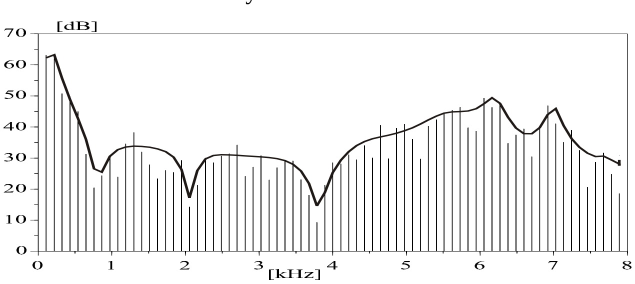

Fig. 7. Analyzed period of /z/ by 20 poles and 10 zeros.

subsequent analysis of voiced speech periods, the order of the analysis cannot be adapted automatically to the corresponding speech segments, so that a fixed number of poles and zeros for all sounds is preferable, especially during the transition of one sound to another. Figure 8 shows that the procedure can be applied to non-periodic signals, too. A segment of the fricative /ʃ/ is analyzed with 20 poles and 10 zeros after a von Hann window has to been applied to the signal. All speech examples are recorded at a sampling rate of 16 kHz and the number of iterations for the algorithm was about 10.

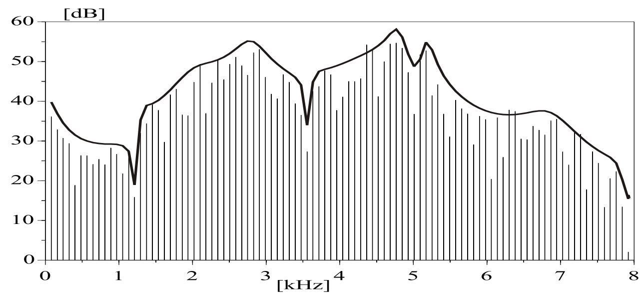

Fig. 8. Analyzed segment of /ʃ/ by 20 poles and 10 zeros.

4. SUMMARY

The proposed algorithm is able to minimize the power of the error signal of a general pole zero system by the use of iterations based on partial solutions. Therefore the algorithm is able to estimate poles and zeros of linear systems from time signals. To study the performance of the procedure analyses of test signals are carried out, which yield a perfect match of the DFT spectra of these signals with the estimated magnitude responses. Examples of analyzed speech signals show, that the procedure is also able to approximate the DFT spectra of speech signals adequately especially with respect to gaps in the speech spectrum.

5. REFERENCES

[1] Burg, J.: “A new analysis technique for time series data”, NATO Advanced Study Institute on Signal Processing, Enschede 1968.

[2] Prony, R.: “Essai éxperimental et analytique ... ”, Journal de l’École Polytechnique ou Bulletin du Travail fait a cette École, 2. Cahier, Paris an IV (1795) pp. 24-76.

[3] Kumaresan, R, Tufts, D.W.: “Estimation the Parameters of Exponentially Damped Sinusoids and Pole-Zero Modeling in Noise”, IEEE Trans. ASSP-30 (1982), pp. 833-840 .

[4] Steiglitz, K.: “On the Simultaneous Estimation of Poles and Zeros in Speech Analysis”, IEEE Trans. ASSP-25 (1977), pp. 229-234.

[5] Song, K.H.; Un, C. K.: “On Pole-Zero Modelling of Speech”, Proc. ICASSP-80, pp. 162-165.

[6] Broersen P.M.T.: “Accurate ARMA Models with Durbin’s Second Method”, Proc. ICASSP-99 (CD-ROM).

[7] Schnell, K.; Lacroix A.: “Parameterbestimmung für Pol-Nullstellen-Modelle”, 10th Conf.: ‘Elektronische Sprachsignalverarbeitung' ESSV-99, Görlitz (Germany) 1999, pp. 64-71.

[8] den Brinker, A.C.; Oomen A.W.J.: “Fast ARMA modelling of power spectral density functions”, Proc. EUSIPCO-2000, Tampere, pp. 1229-1232.