GENERALIZED LINEAR PREDICTION BASED ON ANALYTIC SIGNALS

Unto K. Laine

Helsinki University of Technology
Laboratory of Acoustics and Audio Signal Processing
Otakaari 5A, FIN-02150 Espoo, Finland
E-mail: Unto.Laine@hut.fi

ABSTRACT

The conventional theory of linear prediction (LP) is renewed and extended to form a more flexible algorithm called generalized linear prediction (GLP). There are three new levels of generalization available: On the first level (I) the predictor FIR is replaced with a generalized FIR constructed out of allpass sections having complex coefficients. On the second level (II) the allpass filters have distributed coefficients, i.e., they are unequal, and on the third and the most general level (III) the filter sections may have different characteristics.

The theory of GLP is presented and the algorithm is tested with speech signals. The results show that GLP works as desired: nonuniform frequency resolution can be achieved and the resolution is controlled by the choice of the allpass parameters. On level I, the angle of the pole-zero-pair of the allpass sections defines the highest resolution area while the radius of the pole controls the degree of the resolution improvement. The GLP prediction error decreases rapidly with the order of the predictor. Its normalized RMS value falls off exponentially and its spectral flatness improves efficiently. On the average the results are clearly better than those of conventional LP. Levels II and III are only briely discussed.

1. INTRODUCTION

Linear prediction (LP) is widely applied in speech analysis and coding. Strube [1] has proposed a frequency warped version of conventional LP. The warping is realized by replacing the unit delays of the predictor with first order allpass sections. The warping leads to nonuniform frequency resolution. The idea of frequency scale warping was introduced by Oppenheim, Johnson and Steiglitz [3, 4] in the early 70's. Before that similar structures were applied by Lee [5] to produce orthonormal bases like Laguerre, Fourier, and Legendre for frequency domain filter design. Recently warped LP was further examined by Laine et al. [2], where the warping was treated based on the orthogonal FAM- and FAMlet-transforms.

Only two types of warping can be realized by the first order allpass sections: when the pole of the allpass filter is on the positive real axis the highest frequency resolution point is located at zero frequency and, when the pole is on the negative real axis the highest resolution point is at the folding frequency. In other words, within this formulation the high resolution point can not be placed at any frequency of interest. This problem was also discussed by Oppenheim who proposed the use of allpass sections having complex coefficients. However, the idea was not analyzed further in his paper and even up to now there have been very few publications about this formulation.

In this paper the first order allpass sections with complex coefficients are applied to the autocorrelation LP to produce a freely movable high resolution point at any desired frequency. This leads to a predictor filter which is of a generalized FIR type and which handles complex valued analytic signals. The new LP formulation is called generalized linear prediction (GLP).

2. GENERALIZED LINEAR PREDICTION

First order allpass sections with complex coefficients have an average group delay of one sample and their transfer function consists of a single complex pole inside the unit circle located at the angle θ and radius r from the origin ($r<1$). Each section also has a single complex zero at the same angle and radius $1/r$ from the origin. With a real input signal these filters produce complex valued outputs, i.e., analytic signals. Frequency warping is controlled by the locations of the pole and the zero: the closer they are to the unit circle the stronger is the warping, and, the angle of these special points control the frequency of the largest warping (which is equal to the frequency of the highest resolution). In the following, conventional LP is modified in order to solve for the complex valued predictor coefficients. The real part of the predictor output is then used as a predicted estimate for the real input signal.

Let $A_k(z)$ denote an analytic transfer function (with complex valued coefficients) of the k^{th} first order allpass filter section of the predictor. In level II generalization they all may be different. In the following the most simple case (level I) where the sections are identical is mainly considered. The transfer function of the generalized predictor is given by (1):

$$P(z) = \sum_{k=1}^{p} \gamma_k \prod_{l=1}^{k} A_l(z) \qquad (1)$$

In (1) predictor coefficients γ_k are complex valued and p denotes the total number of filter sections. Now, let $x(n)$ denote the incoming real valued signal and $r_k(n)$ the real part and $i_k(n)$ the imaginary part of the k^{th} section output so that the output of the section $A_k(z)$ can be given in the form $y_k(n) = r_k(n) + j\, i_k(n)$. The predicted value $x_{pre}(n)$ is given as the real part of the linear combination of the tap outputs (2).

$$x_{pre}(n) = \text{Re}[\sum_k \gamma_k\, y_k] = \sum_{k=1}^{p} \alpha_k\, r_k(n) - \sum_{k=1}^{p} \beta_k\, i_k(n) \qquad (2)$$
$$\gamma_k = \alpha_k + j\,\beta_k$$

The mean squared error to be minimized is:

$$E_n = \sum_n \left[x(n) - x_{pre}(n)\right]^2$$
$$= \sum_n \left[x(n) - \sum_{k=1}^{p} \alpha_k\, r_k(n) + \sum_{k=1}^{p} \beta_k\, i_k(n)\right]^2 \qquad (3)$$

To solve for the optimal coefficients γ_k we first expand (3) and then set the partial derivatives equal to zero.

0-7803-2431-5/95 $4.00 © 1995 IEEE

$$E_n = \sum_n [\, x^2(n) + \left(\sum_{k=1}^{p} \alpha_k \, r_k(n) \right)^2 + \left(\sum_{k=1}^{p} \beta_k \, i_k(n) \right)^2 -$$

$$2\, x(n) \sum_{k=1}^{p} \alpha_k \, r_k(n) + 2\, x(n) \sum_{k=1}^{p} \beta_k \, i_k(n) - \qquad (4)$$

$$2 \sum_{k=1}^{p} \alpha_k \, r_k(n) \sum_{k=1}^{p} \beta_k \, i_k(n) \,]$$

The partial derivatives of the t^{th} alpha give:

$$\frac{\delta E_n}{\delta \alpha_t} = \sum_n [\, 2\, r_t(n) \sum_{k=1}^{p} \alpha_k \, r_k(n) - 2\, x(n)\, r_t(n) - $$

$$2\, r_t(n) \sum_{k=1}^{p} \beta_k \, i_k(n) \,] = 0 \qquad (5)$$

By changing the order of the summations and reordering the terms we get :

$$\sum_{k=1}^{p} \alpha_k \sum_n r_t(n)\, r_k(n) - \sum_{k=1}^{p} \beta_k \sum_n r_t(n)\, i_k(n) = \sum_n x(n)\, r_t(n)$$

$$t = 1, 2, \ldots, p \qquad (6)$$

Note that if the signals at section outputs are purely real the correlation terms between real and imaginary parts equal zero and (6) reduces to the normal equations [6]. A similar set of equations is obtained by differentiating (3) with respect to coefficients β_t.

$$\sum_{k=1}^{p} \alpha_k \sum_n i_t(n)\, r_k(n) - \sum_{k=1}^{p} \beta_k \sum_n i_t(n)\, i_k(n) = \sum_n x(n)\, i_t(n)$$

$$t = 1, 2, \ldots, p \qquad (7)$$

In order to solve for complex coefficients γ_k we have to solve (6) and (7) simultaneously. For further derivation we rewrite them in matrix notation.

$$\begin{aligned} C_{rr}\, \overline{\alpha} - C_{ri}\, \overline{\beta} &= \overline{c}_{xr} \\ C_{ri}^T\, \overline{\alpha} - C_{ii}\, \overline{\beta} &= \overline{c}_{xi} \end{aligned} \qquad (8)$$

In (8) the correlation terms $c_{rr}(t,k)$ of the matrix C_{rr} are formed by correlating the real signals of the t^{th} and k^{th} tap outputs. Correspondingly the elements $c_{ri}(t,k)$ of the matrix C_{ri} are formed by correlating the real signal of the t^{th} tap output with the imaginary signal of the k^{th} tap output. Elements $c_{ii}(t,k)$ are obtained by correlating the imaginary signals of the corresponding outputs. The elements of correlation vectors $\boldsymbol{c}_{xr}$ and $\boldsymbol{c}_{xi}$ are solved by correlating the incoming sequence with the real and imaginary outputs of the tap sections. Finally, the unknown alphas and betas are solved from (8).

$$\begin{aligned} \overline{\alpha} &= \left[C_{rr} - C_{ri}\, C_{ii}^{-1}\, C_{ri}^T \right]^{-1} \left(\overline{c}_{xr} - C_{ri}\, C_{ii}^{-1}\, \overline{c}_{xi} \right) \\ \overline{\beta} &= \left[C_{ii} - C_{ri}^T\, C_{rr}^{-1}\, C_{ri} \right]^{-1} \left(C_{ri}^T\, C_{rr}^{-1}\, \overline{c}_{xr} - \overline{c}_{xi} \right) \end{aligned} \qquad (9)$$

3. SOME FEATURES OF GLP

From equations (9) it is possible to produce three levels of generalization: On the first level (I) the predictor FIR is replaced with a generalized FIR constructed of allpass sections having complex coefficients. On the second level (II) the allpass filters have distributed coefficients, i.e., they are unequal, and on the third and the most general level (III) the filter sections may have different characteristics.

Up to now only level I has been studied in detail. The second has been explored in some cases and the third only in a few special cases. On all these levels GLP will produce pole-zero (ARMA) models for the signal. On level I the model is limited in the sense that all the poles of the predictor (or all the zeroes of the spectral model) are located at the same point in the unit circle (multiple pole/zero). In the cases of level II and III more complicated models can be produced.

A general feature of GLP is that when the resolution is improved around some frequency area it is simultaneously lowered at some other points. In other words, *the average resolution is not changed*, but rather the *distribution* of the resolution along the frequency axis.

4. EXAMPLES OF GLP ANALYSIS

The level I GLP algorithm was programmed with *Mathematica* and was used to analyze Finnish vowel sounds. The vowels were sampled at 22.25 kHz and quantized to eight bits. The pre-emphasized incoming signal was windowed and filtered by the cascaded allpass sections (the generalized FIR structure). The components of the correlation matrices were then produced by forming dot products between the analytic tap output signals and the real incoming one.

The results of the GLP analysis were compared to those given by classical LP. On the average GLP needs half the order of LP because the predictor coefficients are complex valued. In GLP case the results of the analysis depend not only on the order of the predictor but also on the allpass coefficients used. This new freedom introduces a new problem: how to choose and optimize the coefficients? In this study the normalized energy of the error signal and the spectral flatness of the same were used to form the criteria. However, in the GLP case the flattness is not equal over the frequencies but is *frequency dependent.* In those areas where the resolution is highest the flatness reaches its maximum too. In other areas the flatness may be even less than in the corresponding LP case. In the following, two cases are considered in more detail: vowels /oe/ and /y/ (vowel /y/ may be considered as a rounded variant of the English vowel /i/).

4.1 Normalized residual RMS value

The most surprising result of the simulations of the GLP algorithm is related to the behaviour of the normalized RMS value of the GLP residual. In the LP case this decreases more slowly when the order of the predictor increases, whereas the logarithmic RMS value of the GLP residual continues to decrease *linearly* (see Fig. 1). Finally, the value approaches the limit of numeric underflow. With double precision computation (with about 18 decimals) this point was reached with order 10-13 depending on the parameters of the allpass sections. Since the spectral model fits well to the Fourier spectrum up to the underflow point and the flatness improves too, the behaviour of the residual energy should be a true GLP related phenomenon and not related to any artifact.

In Fig. 1 the /oe/ vowel was analyzed with allpass parameters: $\theta = 48*\pi/256$. (corresponds to 2.09 kHz), r = 0.7. The 0th order predictor means no predictor at all, and the resulting error is equal to the incoming signal. Thus the normalized RMS error is one, i.e., 0 dB. When the predictor has an order higher than four the normalized RMS error will be divided by 3.236 (-10.2 dB on the log scale) when the order of the predictor increases by one. The corresponding RMS error of conventional LP decreses very slowly as seen in the figure. Further investigations are needed to reveal the reasons for this large difference between the methods.

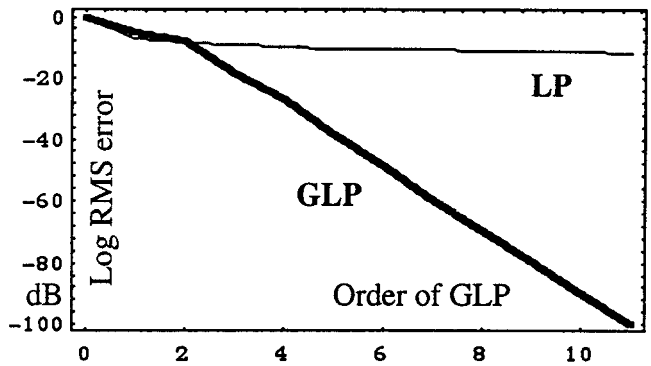

Fig. 1. Log RMS value of the normalized GLP residual (thick line) as a function of predictor order (0 - 11) compared with the corresponding LP residual value (thin line). Note that the x-axis denotes the actual LP order (0 - 22) divided by two.

4.2 Derived model and spectral flatness

The derived GLP model is fitted to the Fourier spectra of the vowel /oe/ in Fig. 2. The chosen allpass parameters lead to a high accuracy fit around 2.1 kHz whereas around 8-10 kHz the fit is not the best possible. Due to the nonuniform resolution there is a increasing mismatch towards the higher frequencies. This difference can be corrected by a simple fixed filter the shape of which depends on the actual allpass coefficients.

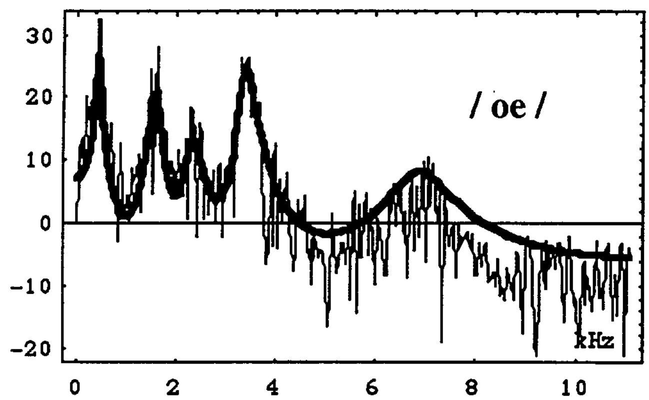

Fig. 2. Fourier spectrum of the pre-emphasized vowel /oe/ and the derived GLP model of order 11.

In our case the model should have less energy at higher frequencies. Correspondingly the derived residual should be emphasized in the high frequency area. This is clearly seen in Fig. 3, where the spectra of the derived residual is shown. The correction needed is about 10 dB.

The spectral flatness of the residual signal was computed over eight frequency bands (from DC to n•1.39 kHz, n=1, 2,..., 8) in order to investigate the nonuniform frequency resolution produced by GLP. Fig. 4 shows the derived result in the case of the vowel /oe/. When the whole band (11.1 kHz) is analyzed we get the classical flatness measure. The increasing flatness indicates succeeded modeling and decreasing that the modeling is in trouble there. In Fig. 4 it can be seen that LP has problems to model the spectra (decreasing curve) around 3 - 4 kHz, where the fourth formant is located. Both GLP flatness curves run clearly higher from DC up to 5 - 8 kHz even though the order of the GLP is less than half of that of the LP (7 vs. 22). The tilting of the produced residual spectra will cause an additional error in the higher frequencies (decreasing curve). However, when this tilting is compensated for (GLP-7c) the GLP model is comparable with the LP model even in the high frequency area. When the vowels are analyzed for coding purposes it is very important that the model fits well in the frequency band where the most important formants are located and also where the selectivity of the human auditory system is high. We can note that these requirements are well achieved by GLP-7.

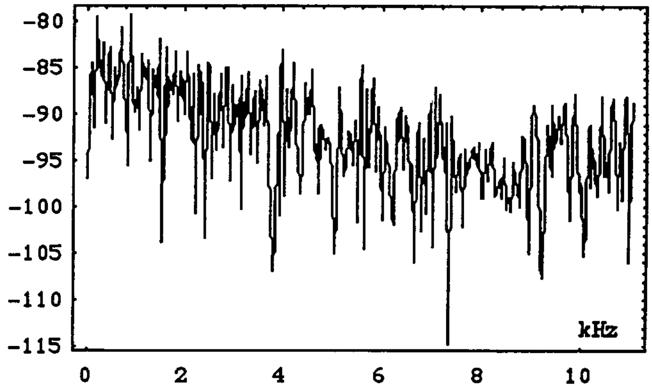

Fig. 3. Spectrum of the GLP residual.

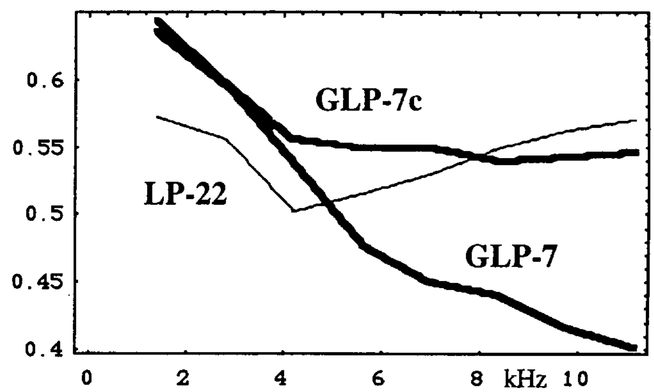

Fig. 4. Spectral flatness of the GLP and LP residual measured over eight bands. Vowel /oe/, GLP-7 (thick lines) and LP-22 (thin line) compared. In the case GLP-7c the spectral tilting of the residual has been corrected.

Fig. 5 shows the GLP-11 model and the corresponding spectra of the /y/ vowel. In this case the allpass parameters were changed to: $\theta = 32*\pi /256$. (corresponds to 1.39 kHz), $r = 0.7$. Also here the tilting of the model (and the spectrum of the residual) is clearly seen. An interesting detail is that now the model even picks up an additional peak between formants one and two. This additional peak may be due to a subglottal resonance. Fig. 6 gives a closer picture about this area. This type of weak top between two stong formants is extremely seldom modeled by conventional LP. Due to the fact that the GLP resolution is improved around just these frequencies it was able to find this peak.

In the case of the vowel /y/ the spectral flatness is clearly better with GLP-11 than with LP-22 as seen in Fig. 7. Also here the spectral tilting will produce an additional error, which can be compensated for with a proper filter as discussed above (compare cases GLP-11c and GLP-11 in Fig. 7).

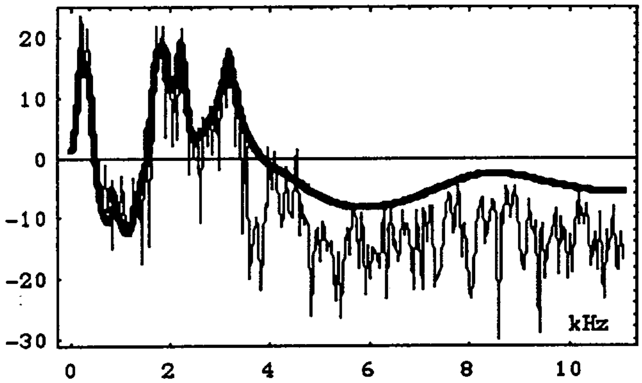

Fig. 5. Fourier spectrum of the pre-emphasized vowel /y/ and the derived GLP model of order 11.

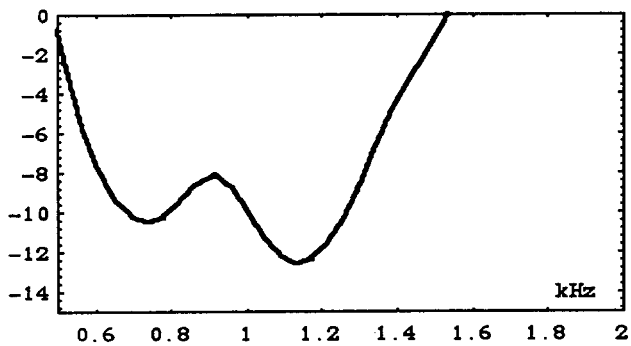

Fig. 6. A possible subglottal resonance at 0.9 kHz modeled by the GLP model.

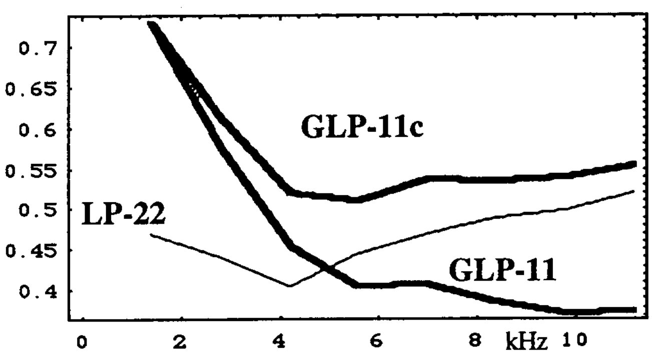

Fig. 7. Spectral flatness measured over eight frequency bands. Vowel /y/ analyzed with GLP-11 and LP-22.

4.3 Modeling of spectral details

Dealing with the nonuniform spectral resolution of the GLP, a natural question is: can it easily model spectral details just by focusing the maximum resolution to the frequency of interest. In Fig. 8 the GLP is focused to 3.47 kHz ($\theta = 80\pi/256$, $r = 0.9$) in the analysis of vowel /oe/. The vowel has a strong formant just centerd around that frequency. Fourier spectra shows also some harmonic peaks amplified by the formant. As demonstrated by Fig. 8 the GLP model picks up the strongest harmonics at the formant. Outside of this high resolution area GLP models only approximately the formant locations and further apart from the high resolution locus the model does not give any information about the spectral variation (flat response). However, the general shape of the spectra is still described by the model. Estimates for positions of the harmonics can be solved from the GLP model. In this example the frequency difference between the neigthboring harmonics are between 115 - 127 Hz. In order to get better estimates, the model should focus iteratively further to the individual peaks.

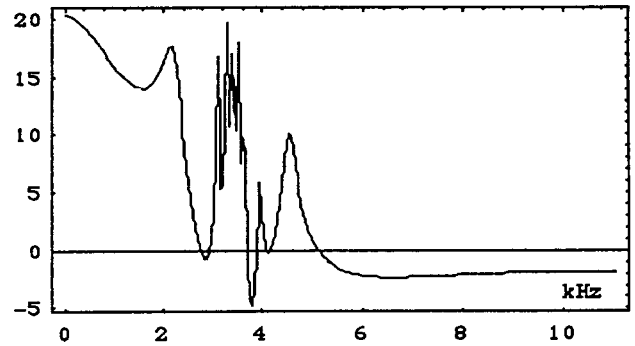

Fig. 8. GLP-11 model focused to the fourth formant of the /oe/ vowel.

5. DISCUSSION

Generalized linear prediction, as derived and experimented with in this paper, has proven to own many new useful and interesting properties. An interesting novelty is that it is able to model the spectral shape *locally* while at the same time giving just a hint about the global behaviour of the spectrum. We are able to imagine many new interesting applications for the GLP method. In this preliminary report we only weakly tuched on some of those areas. Future experimentation with level II and III GLP methods will give even more tools for linear prediction based signal processing.

ACKNOWLEDGMENT

This study has been financed by the Academy of Finland.

REFERENCES

[1] Strube H. W., "Linear prediction on a warped frequency scale", J. Acoust. Soc. Am., 64 (4), pp. 1071-1076, Oct. 1980.
[2] Laine U. K., Karjalainen M., and Altosaar T., "Warped linear prediction (WLP) in speech and audio processing", Proc. ICASSP-94, Adelaide, South Australia, III pp. 349-352 , 1994.
[3] Oppenheim A. V., Johnson D. H., Steiglitz K., "Computation of spectra with unequal resolution using the fast Fourier fransform", Proc. IEEE, 59, pp. 299-301, Feb. 1971.
[4] Oppenheim A. V., Johnson D. H., "Discrete representation of signals", Proc. IEEE, 60, No. 6, pp. 681-691, June 1972.
[5] Lee Y. W., *Statistical Theory of Communication*, ch. 19: "Synthesis of optimum linear systems by means of orthonormal functions", Wiley, New York, 1960.
[6] Makhoul J., "Linear Prediction: A Tutorial Review," Proc. IEEE, 63, pp. 561-580, 1975.