Lattice LP Filtering for Noise Reduction in Speech Signals

Erhard Rank, Gernot Kubin

Signal Processing and Speech Communication Laboratory, Graz University of Technology
Inffeldgasse 12, A-6010 Graz, Austria
erank@tugraz.at

10.21437/Interspeech.2006-73

Abstract

We present a simple yet effective algorithm for noise reduction of speech signals using a lattice LP filter. Based on previous investigations and a theoretical analysis of the lattice filter parameter estimation we introduce an improved parameter estimation algorithm that takes into account the non-stationary nature of speech and expected noise signals, yielding a good suppression of stationary and slowly time-varying noise. The algorithm has zero delay for the speech signal, promoting its application for telephony or hearing aids. No additional or explicit noise estimation algorithm is needed.

Index Terms: noise reduction, speech recognition.

1. Introduction

Noise reduction for speech signals is a task becoming more and more important in telephony, for automatic speech recognition, but also for applications, like, e. g., digital hearing aids. The challenges we are facing are non-white, non-stationary or highly impulsive noise.

Linear prediction (LP) [1] is an important tool for speech processing and is applied in numerous applications for speech transmission, analysis, recognition, and synthesis. The implementation of LP filters can be of direct form, or using a lattice structure [2, 3], which have been used for noise reduction, e. g., in [4, 5].

In this paper we present a lattice filter LP predictor and its application to noise reduction, starting with a short review of linear prediction and the lattice filter in sect. 2, the principle of its application to noise reduction in sect. 3, and a motivation of the corrections to the standard parameter estimation method based on estimates of the reflection coefficients from a noisy signal in sect. 4. The proposed algorithm is presented , with examples for the denoising of speech signals, in sect. 5, and we finish with conclusions and outlook.

2. Linear prediction and lattice filter

Linear prediction is commonly applied to a speech signal $x(n)$, e. g., to reduce the variance of a speech signal for transmission, by using a low-order, slowly time-varying finite impulse response filter to predict a signal sample:

$$\hat{x}(n) = \sum_{i=1}^{M} b_i(n)\, x(n-i) \quad . \tag{1}$$

Here, M is the order of the LP filter, and $b_i(n)$ are the filter coefficients, which are estimated based on the signal properties and updated on a frame-wise basis, e. g., each 10 ms. Algorithms directly providing coefficients $b_i(n)$ for the direct form FIR filter in (1) are the so-called "auto-correlation method" or the "covariance method." LP for speech signals frequently uses a low filter order ($M = 10 \dots 20$, depending on the sampling rate) to model the spectral envelope.

An prediction filter equivalent to the direct form filter is the lattice filter [2, 3], which has a direct relation to a physical model of the vocal tract [1]. The lattice prediction filter operation is characterized by the equations

$$f_0(n) = b_0(n) = x(n) \ ,$$

$$f_m(n) = f_{m-1}(n) + k_m(n)\, b_{m-1}(n-1) \ , \tag{2}$$

$$b_m(n) = b_{m-1}(n-1) + k_m(n)\, f_{m-1}(n) \ , \tag{3}$$

evaluated at each time n for all lattice stages $m = 1 \dots M$. $f_m(n)$ and $b_m(n)$ are the forward and backward error in stage m at time n, respectively, and $k_m(n)$ are the reflection coefficients of the lattice filter (in a more general formulation of the lattice filter the reflection coefficients in (2) and (3) are individual different parameters, however, we will use the formulation with equal 'forward' and 'backward' reflection coefficients here). A schematic of this lattice filter is given in fig. 1.

The forward error at stage M is the prediction error signal of the lattice LP filter:

$$f_M(n) = e(n) = x(n) - \hat{x}(n) \ . \tag{4}$$

Optimal reflection coefficients k_m for minimizing the mean squared prediction error of an undistorted signal are found by

$$k_m(n) = -\frac{r_{m-1}(n)}{q_{m-1}(n)} \ , \tag{5}$$

with the expected values for the forward and backward error correlation and power

$$r_m(n) = E\{f_m(n)\, b_m(n-1)\} \ , \tag{6}$$

$$q_m(n) = \frac{1}{2} E\{f_m^2(n) + b_m^2(n-1)\} \ . \tag{7}$$

Commonly, the expectation operators E in (6) and (7) are evaluated using low-pass filtered instantaneous values of $f_m(n)\, b_m(n-1)$ and $f_m^2(n) + b_m^2(n-1)$, respectively, for example using one-pole recursive low-pass filters (lossy integration, cf. sect. 5). Hence, the adaption of the lattice filter to model the slowly time-varying input speech signal is done by evaluating (6), (7), and (5) for each time n after filter operation, (2) and (3) – as opposed to the frame-wise coefficient update of the direct form LP filter in (1).

We kindly acknowledge the support by the European Union for the FP6 IST STREP SNOW (Services for NOmadic Workers, FP6-511587).

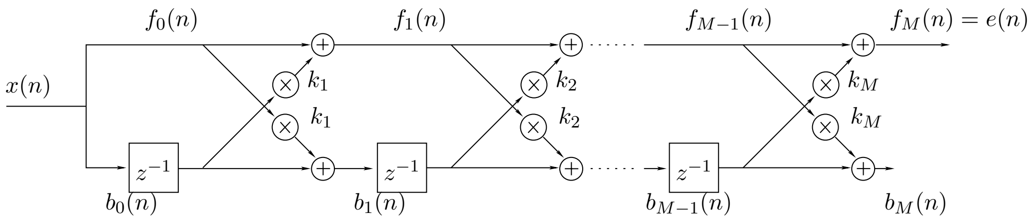


Figure 1: Lattice filter

3. Noise reduction

For the noise reduction task we consider an observed signal $y(n)$ originating from a linear additive noise model

$$y(n) = x(n) + \varepsilon(n) \ , \tag{8}$$

with the speech signal component $x(n)$ and an additive background noise component $\varepsilon(n)$. The task of noise reduction is to provide a good estimate for the speech signal component $x(n)$. In the single-channel setting considered here (and for the SNOW project), this estimate is based on the noisy signal observation $y(n)$ only and does not make use of additional information (like, e. g., a second signal from a microphone recording only background noise).

Reduction of background noise in speech signals using linear prediction filtering can be based on the assumption that the speech signal component is well predicted whereas the noise component is not. Thus, the predicted signal $\hat{x}(n)$ may be taken as an estimate for the speech component. While for the direct form prediction filter in (1) the output is the predicted signal immediately, for the lattice filter, $\hat{x}(n)$ is computed effectively as the difference between input signal and output of the forward prediction path:

$$\hat{x}(n) = y(n) - e(n) \tag{9}$$

for the speech component estimate, with $e(n) = f_M(n)$, cf. (4).

In the application of a lattice LP filter to noise reduction in [4]—where a high order filter ($N = 256$) is used to enable the modeling of the spectral fine structure of speech—it is observed that $r_m(n)$ exhibits a large variance due to the noise signal component in the higher filter stages. It is suggested to reduce the according variance of the reflection coefficients by using a fixed (large) value for the power estimates $q_m(n) = G$. The reduction of the magnitude of reflection coefficients, or the reduction of the radii of zeros of the LP filter transfer function, has been proposed also for other purposes, like, e. g., better modeling of the spectral envelope or more accurate formant estimation.

In the next section we will motivate a reduction of the magnitude of the reflection coefficients by the derivation of minimum mean-square error estimators for r_m and q_m.

4. Estimator for reflection coefficients from the noisy signals

For the computation of reflection coefficients for the lattice filter, or of the partial correlations (PARCORs), which are equal to $-k_m$, based on estimates from a noisy signal, we assume whiteness for the additive noise signal $\varepsilon(n) \in \mathcal{N}(0, \sigma_n^2)$, which should also be uncorrelated with $x(n)$. This constitutes the least informed (maximum entropy) model.

We are aware that this assumption is not realistic for environmental noise in general.However, as we will show, it gives evidence of the need for a correction of the reflection coefficients.

The estimates for correlation (6) and power (7) for the computation of the reflection coefficients in (5) are now based on the noisy observed signal $y(n)$, and we will show the need for a corrective term to yield estimates for reflection coefficients $\hat{k}_m$ related to the noise-free signal $x(n)$.

In particular, for the estimation of the reflection coefficient in the first filter stage $m = 1$, we get the following expectation for r_0:

$$\begin{aligned} r_0 &= E\{f_0(n)\, b_0(n-1)\} \\ &= E\{y(n)\, y(n-1)\} \\ &= E\{(x(n) + \varepsilon(n))(x(n-1) + \varepsilon(n-1))\} \\ &= E\{x(n)\, x(n-1)\} \ . \end{aligned} \tag{10}$$

For the error power estimate q_0 in the first filter stage we get

$$\begin{aligned} q_0 &= \frac{1}{2} E\{f_0^2(n) + b_0^2(n-1)\} \\ &= \frac{1}{2} E\{y^2(n) + y^2(n-1)\} \\ &= \frac{1}{2} E\{(x(n) + \varepsilon(n))^2 + (x(n-1) + \varepsilon(n-1))^2\} \\ &= \frac{1}{2} E\{x^2(n) + x^2(n-1)\} + \sigma_n^2 \ . \end{aligned} \tag{11}$$

The resulting error in the values of the reflection coefficients is depicted in fig. 2,

Thus, to come up with reflection coefficients related to the noise-free signal $x(n)$, the correlation estimate from the noisy observation can be used unchanged $\hat{r}_0 = r_0$, whereas the error power estimate has to be corrected as

$$\hat{q}_0 = q_0 - \sigma_n^2 \ , \tag{12}$$

and the corrected reflection coefficient is computed as

$$\hat{k}_1 = -\frac{\hat{r}_0}{\hat{q}_0} = -\frac{r_0}{q_0 - \sigma_n^2} \ . \tag{13}$$

Introducing $\gamma = \frac{E\{y^2\}}{\sigma_n^2}$ (where $\gamma - 1 = \frac{E\{y^2\} - \sigma_n^2}{\sigma_n^2}$ is the a posteriori SNR), and noting that $q_0 = \frac{1}{2} E\{f_0^2(n) + b_0^2(n-1)\} = E\{y^2\}$, we can rewrite this equation to

$$\hat{k}_1 = -\frac{1}{1 - \frac{\sigma_n^2}{q_0}} \frac{r_0}{q_0} = \frac{1}{1 - \frac{1}{\gamma}}\, k_1 \ . \tag{14}$$

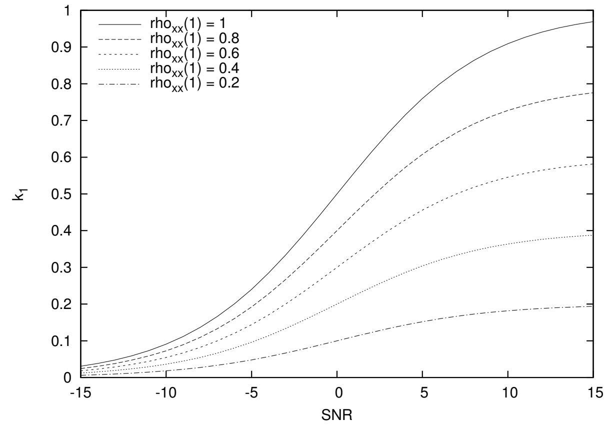


Figure 2: Values of the reflection coefficient k_1 as computed from the noisy signal without correction as a function of a priori SNR for different values of the auto-correlation $\rho_{xx}(1)$ of the noise-free signal $x(n)$

This means a scaling of the reflection coefficient k_1 as originally computed for the distorted signal $y(n)$ using (5), (6), and (7) by a factor $\frac{1}{1-\frac{1}{\gamma}}$.

Equation (14) can be generalized for the higher lattice stages $m = 2, 3, \ldots$, which yields a recursive correction of the reflection coefficients $\hat{k}_m$. In our experiments directly applying the corrections deduced above, however, severe problems (like intermediate instabilities of the resulting LP analysis (!) filter) have been observed[1].

Nevertheless, from the above we can infer that a modification of reflection coefficients in magnitude, i. e., a modification of the ratio between correlation and power estimate is beneficial for the prediction of a signal $x(n)$ when observing a signal $y(n)$ containing additive noise. Finding the correction terms, however, relies on good estimates for the signal and noise power, σ_x^2 and σ_n^2, respectively. Furthermore, the model up to now does not take into account any knowledge about properties of speech and of the expected noise signal. We will now present a method to yield a correction of reflection coefficients based on simple assumptions about the variation of the correlation and power of the speech and noise signals over time.

5. Correlation and power estimates for non-stationary noisy signals

As noted above, the estimates for error correlation (6) and error variance (7) are commonly based on low-pass filtering of the instantaneous values. Often one-pole low-pass filtering (lossy integration) is used:

$$\tilde{r}_m(n) = \lambda_r \, \tilde{r}_m(n-1) + f_m(n) \, b_m(n-1) \ , \qquad (15)$$

$$\tilde{q}_m(n) = \lambda_q \, \tilde{q}_m(n-1) + \frac{1}{2}(f_m^2(n) + b_m^2(n-1)) \ , \qquad (16)$$

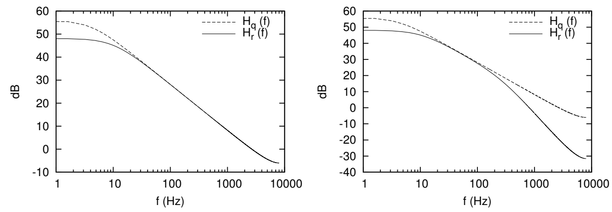
[1]We used simple signal and noise estimates based on $q_0(n)$ and its minimum statistics, respectively, or a noise estimation from $e(n)$, which, however, forms a feedback system. Instabilities were observed in both cases.

with the same poles (forgetting factors) $\lambda_r = \lambda_q$ for both correlation and power estimate.

Our proposal is to allow for different pole positions $\lambda_q \geq \lambda_r$. The resulting filter transfer functions

$$H_r(z) = \frac{1}{1 - \lambda_r z^{-1}} \ , \qquad H_q(z) = \frac{1}{1 - \lambda_q z^{-1}} \qquad (17)$$

for $\lambda_r = 0.99608$ and $\lambda_q = 0.99843$ and a sampling rate of 16 kHz are depicted in fig. 3 (a). As can be seen the ratio of $\tilde{r}_m(n)$ and $\tilde{q}_m(n)$ will be influenced less for low frequencies, i. e., for slowly varying correlation and power, whereas for faster variations (above ≈ 10 Hz) the ratio is unchanged as compared to estimates with $\lambda_r = \lambda_q$. Under the assumption that these parameters vary faster for the speech signal (considering, e. g., a phoneme rate of 10 per second) than for the noise signal (stationary or slowly time varying noise), the resulting lattice prediction filter will predict the speech signal component well, whereas the noise component is suppressed.



(a) (b)

Figure 3: Frequency characteristic of the low-pass filters for error correlation $H_r(z)$ (solid lines) and variance $H_q(z)$ (dashed lines), for (a) two one-pole low-pass filters with $\lambda_r = 0.99608$ and $\lambda_q = 0.99843$, (b) a one-pole low-pass for the power estimate $\tilde{q}(n)$ with $\lambda_q = 0.99843$, and a two pole low-pass for the correlation estimate $\tilde{r}(n)$ with $\lambda_{r1} = 0.99608$ and $\lambda_{r2} = 0.9$. The larger the distance between the two transfer functions the more noise suppression occurs.

To address impulse noise, we propose to reduce the ratio between correlation and power estimate also for high frequencies, which can be done by applying a second pole in the low-pass filter for the correlation $H_r(z)$. An according example transfer function is depicted in fig. 3 (b).

The LP filter order M necessary to achieve good noise reduction may be chosen astonishingly low, even lower than the order commonly used for modeling the spectral envelope of speech signals. For the example in fig. 4 a predictor with order $M = 10$ was used for a signal with a sampling rate of 16 kHz. This example contains several occurrences of strongly non-stationary noise bursts, which are well removed by our algorithm. The signal is an example recording for the European project SNOW (Services for NOmadic Workers)—in the scope of which this algorithm has been developed—from a factory floor environment, i. e., an adverse acoustical environment.

First informal tests of this algorithm in combination with and as a replacement for the ETSI advanced front-end [9] in the scope of automatic speech recognition show encouraging results, the optimization of parameters and detailed tests, however, still have to be performed.

Computational complexity of the proposed algorithm depend directly on the filter order chosen, and for larger filter order may

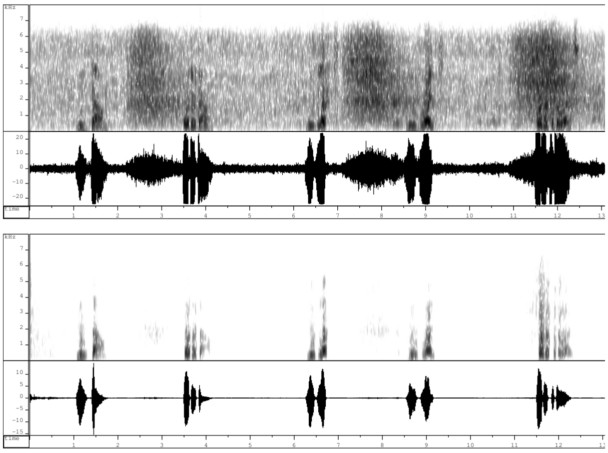


Figure 4: Example for the denoising of a wide-band speech signal

exceed the computational complexity of other (e. g., FFT-based) noise reduction methods, due to the sample-wise processing and parameter update. Note, however, that the sample-wise processing of our algorithm enables the implementation of noise reduction *without delay* of the speech signal, which is a benefit, e. g., for the application in hearing aids.

6. Conclusions and outlook

A simple yet effective algorithm for noise reduction of speech signals using a lattice LP filter has been presented, motivated by the derivation of estimates for the error correlation and power in the lattice filter. In addition, we consider the specific properties of time-variations in speech signals and the expected additive noise. The resulting algorithm is well suited for the removal of stationary and slowly time-varying noise. For the application of this algorithm no noise estimation is required. The algorithm has zero delay of the processed signal, making it applicable for delay-critical noise reduction tasks, such as in hearing aids.

Further investigations will tackle the optimization and testing for the application with automatic speech recognition systems, the promotion of impulse noise suppression capabilities—which are not yet sufficient for the targeted factory floor environment—, as well as the test and optimization for human perception.

7. References

[1] J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech*. Berlin, Heidelberg, New York: Springer, 1976.

[2] J. I. Makhoul and L. K. Cosell, “Adaptive lattice analysis of speech,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, pp. 654–659, June 1981.

[3] M. L. Honig and D. G. Messerschmitt, *Adaptive Filters: Structures, Algorithms, and Applications*. Boston-The Hague-London-Lancaster: Kluwer Academic Publishers, 1984.

[4] A. Kawamura, K. Fujii, Y. Itoh, and Y. Fukui, “A noise reduction method based on linear prediction analysis,” *Electronics and Communications in Japan, Part 3*, vol. 86, no. 3, pp. 1–10, 2003.

[5] M. H. Savoji, “Effective noise reduction of speech signals using adaptive lattice filtering, segmentation and soft decision,” in *IEE Colloquium on New Directions in Adaptive Signal Processing*, pp. 7/1–7/5, Feb. 1993.

[6] G. Kubin, “Nonlinear processing of speech,” in *Speech Coding and Synthesis* (W. B. Kleijn and K. K. Paliwal, eds.), pp. 557–610, Amsterdam: Elsevier, 1995.

[7] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, pp. 1109–1121, Dec. 1984.

[8] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error log-spectral amplitude estimator,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, pp. 443–445, Apr. 1985.

[9] *ETSI ES 202 050 V1.1.3 Speech Processing, Transmission and Quality Aspects (STQ); Distributed speech recognition; Advanced front-end feature extraction algorithm; Compression algorithms*.