Linear Predictive Coding With Backward Adaptation and Noise Shaping

Srikanth Korse, Hauke Krüger, Matthias Pawig, Peter Vary

Institute of Communication Systems and Data Processing, 52074 Aachen
Email: {korse,krueger,pawig,vary}@ind.rwth-aachen.de
Web: www.ind.rwth-aachen.de

Abstract

In linear predictive coding, the benefit of backward adaptation in comparison to forward adaptation of the prediction coefficients is that it allows to save bit rate since no prediction coefficients need to be transmitted from the encoder to the decoder. Also, backward adaptation yields very low complexity and algorithmic delay. Unfortunately, *noise shaping* techniques which have proven to enable higher perceptual quality in speech and audio coding than approaches without noise shaping have been proposed only for forward adaptation.

In order to overcome this shortcoming, in this paper we propose a novel scheme for noise shaping in linear predictive coding with backward sequential adaptation. The new concept is investigated involving two different methods for the adaptation of the prediction coefficients, the *Gradient Adaptive Lattice* (GAL) and the *Normalized Least Mean Square* (NLMS) algorithm. Measurements of the perceptual speech quality for a database of speech examples based on the well-known *Perceptual Evaluation of Speech Quality* (PESQ) measure show that due to the involved noise shaping a significant increase of the perceived quality can be achieved for speech signals sampled at 8 kHz sampling frequency with bit rates of 40, 32 and 24 kbits/s.

1 Introduction

Most modern speech codecs such as the Adaptive Multi-Rate (AMR) Speech Codec [1] and the Adaptive Multi-Rate Wideband (AMR-WB) Speech Codec [2] are based on linear prediction with forward block adaptation. In that context, short segments of the signal to be encoded are buffered and afterwards analyzed by means of the linear prediction analysis to yield the linear prediction (LP) coefficients. As the input signal is only available in the encoder, the computed LP coefficients must be quantized and afterwards transmitted to the decoder as side information in addition to the quantized LP residual signal. The choice of the length of the LP analysis buffer is always a compromise between bit rate, coding efficiency and algorithmic delay. In the end, it turns out that a buffer length equivalent to approximately 20 ms duration well corresponds to the assumption of the typical duration of short-term stationary segments in speech signals [3].

An alternative to forward adaptation is to determine the LP coefficients sample-by-sample by means of backward sequential adaptation which is often denoted as ADPCM (Adaptive Differential Pulse Code Modulation, [3]). Compared to approaches based on forward adaptation, the backward adaptation has the benefit that the linear prediction coefficients are derived from the *reconstructed* input signal on the encoder and the decoder side. Therefore, no coefficients need to be transmitted from the encoder to the decoder, allowing to save bit rate. Moreover, backward adaptive linear predictive coding principally has an algorithmic delay of only 1 sample whereas forward adaptation yields an algorithmic delay in the order of the duration of the LP analysis buffer (typically around 20 ms).

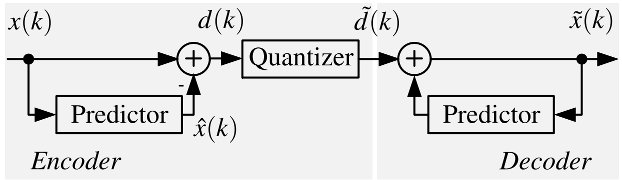


Figure 1: Principle of linear predictive coding (LPC).

Targeting at very low bit rates, it was proven that a spectral shaping of the quantization noise is useful to achieve a higher perceived speech and audio quality. Approaches for noise shaping are well-known for linear predictive coding with forward adaptation of the prediction coefficients, e.g., [4]. However, no technique has been proposed in the past for noise shaping in linear predictive coding with backward adaptation such as, e.g., the ITU-T G.726 codec [5]. In order to solve this issue, a novel approach for noise shaping in linear predictive coding with backward adaptation of the LP coefficients is proposed in this paper.

In the following, the principle of linear predictive coding with forward adaptation of the LP coefficients with noise shaping and the conventional approach involving backward adaptation **without** noise shaping are reviewed in Section 2. Afterwards, the new approach involving backward adaptation of the LP coefficients **and** noise shaping will be introduced in Section 3. Finally, the impact of the proposed noise shaping will be principally reviewed in Section 4 on the basis of a simple test codec. In that section, the new approach will also be evaluated based on the comparison of average PESQ measures obtained for a database of speech signals.

2 Linear Predictive Coding

The basic principle of linear predictive coding (LPC) is shown in Figure 1. The LP analysis filter is part of the encoder and intended to decorrelate the signal to be coded, $x(k)$. It does so by computing an estimate $\hat{x}(k)$ in the so-called linear predictor to approximate the current input sample $x(k)$ by means of an FIR filter of order N_{LPC}, $\hat{x}(k) = \sum_{i=1}^{N_{\mathrm{LPC}}} a_i \cdot x(k-i)$. The output from the predictor is subtracted from the input signal $x(k)$ to yield the prediction residual signal $d(k) = x(k) - \hat{x}(k)$ which is fed into the quantizer. The quantized prediction residual is denoted as $\tilde{d}(k) = \delta(k) + d(k)$ where the impact of the quantizer is commonly modeled as an additive white noise source producing the quantization error $\delta(k)$. $\tilde{d}(k)$ is transmitted to the decoder where $x(k)$ is approximately reconstructed from the signal $\tilde{d}(k)$ by applying the LP synthesis filter which is the inverse of the LP analysis filter. The reconstructed speech signal is denoted as $\tilde{x}(k)$.

In most cases, speech and audio codecs are designed to

ISBN 978-3-8007-3640-9

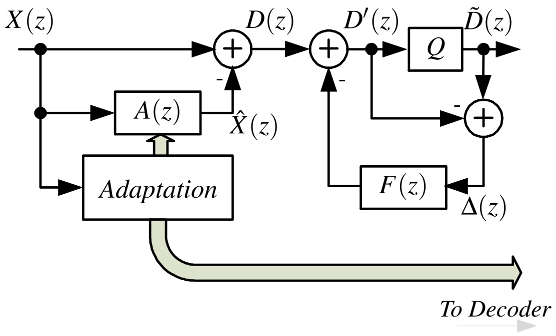


Figure 2: Forward adaptive linear predictive encoder with noise shaping

have similar performance for all types of signals in a wide dynamic range. Therefore, the quantizer is assumed to produce a constant *quantizer signal-to-quantization-noise ratio*,

$$\mathrm{SNR}_0 = \frac{E\{d^2(k)\}}{E\{(d(k)-\tilde{d}(k))^2\}} \tag{1}$$

In order to evaluate the impact of coding in terms of perceived speech or audio quality, the **quantization noise within the reconstructed signal** is $n(k) = x(k) - \tilde{x}(k)$. Correspondingly, the *overall signal-to-quantization-noise ratio* is defined as

$$\mathrm{SNR}_c = \frac{E\{x^2(k)\}}{E\{(x(k)-\tilde{x}(k))^2\}} \tag{2}$$

and typically deviates from SNR_0. Given a stationary input signal $x(k)$, the signals $\delta(k)$ and $n(k)$ can be described by their power spectral densities $\phi_{\delta\delta}(\Omega)$ and $\phi_{nn}(\Omega)$ to better investigate the impact of the noise shaping with Ω as the normalized frequency. It is commonly assumed that the signal $\delta(k)$ is spectrally flat, hence, $\phi_{\delta\delta}(\Omega) = \text{const.}$ The spectral envelope related to the signal $n(k)$ can be influenced by means of noise shaping and will be subject of evaluation in the following.
In order to study the impact of noise shaping, we prefer to use the z-transforms $X(z)$, $D(z)$, $\tilde{D}(z)$, $\tilde{X}(z)$, and $\Delta(z)$ of finite segments (with finite energy) of the signals $x(k)$, $d(k)$, $\tilde{d}(k)$, $\tilde{x}(k)$, and $\delta(k)$, respectively.

2.1 LPC with Forward Adaptation and Noise Shaping (F-LPC)

The specific case of LPC with forward adaptation of the LP coefficients - often denoted as *block adaptive LPC* - and noise shaping is shown in Figure 2. The LP analysis filter is a time variant minimum phase filter given as $1 - A(z)$ with $A(z) = \sum_{i=1}^{N_{\mathrm{LPC}}} a_i \cdot z^{-i}$ of which the LP coefficients a_i are updated frequently. For the *adaptation* of the a_i the input signal is decomposed into (windowed) segments of a specific length. In most cases the auto-correlation method [6] is employed, followed by the well-known Levinson-Durbin algorithm, to derive the LP coefficients based on a criterion to minimize the variance of the prediction residual signal $d(k)$. Signal $\hat{X}(z) = A(z) \cdot X(z)$ from Figure 2 is the output of the predictor and hence the estimate to approximate the input signal $X(z)$ derived from past samples of $X(z)$. The N_{LPC} computed LP coefficients are quantized (e.g., [7]) and employed in the LP analysis filter $A(z)$ to produce the LP residual signal $D(z)$. In order to reconstruct the input signal in the decoder the LP residual signal $D(z)$ is quantized and transmitted to the decoder together with the quantized LP coefficients.
The noise shaping is achieved by feeding back the quantization error $\Delta(z)$ using the filter $F(z)$ which is usually derived from the predictor $A(z)$ as

$$F(z) = A(z/\gamma). \tag{3}$$

The result is signal $D'(z) = D(z) + F(z) \cdot \Delta(z)$ which is fed into the quantizer Q. γ is the noise shaping factor to control the impact of noise shaping as well as the overall quantization SNR to be reviewed more in detail in the following. It is well-known from, e.g., [6] that the quantization noise within the reconstructed signal in the decoder is given as

$$X(z) - \tilde{X}(z) = \Delta(z) \cdot \frac{1-F(z)}{1-A(z)}. \tag{4}$$

Based on this result, the power spectral density of the quantization noise can be expressed as

$$\phi_{nn}^{(\text{F-LPC})}(\Omega) = \phi_{\delta\delta}(\Omega) \cdot \left|\frac{1-F(\Omega)}{1-A(\Omega)}\right|^2. \tag{5}$$

An elaborate evaluation of noise shaping in LPC as well as the introduction of a new *noise propagation model* was recently proposed in [8]. The conclusions shall only be briefly summarized in the following:
The maximum quantization SNR *in case of high bit rates* can be achieved for $\gamma = 1$ and hence $F(z) = A(z)$. In that case the benefit due to the employment of linear prediction analysis and synthesis filtering in terms of overall quantization SNR compared to a direct quantization of signal $x(k)$ can be quantified by means of the *prediction gain* G_p as

$$\mathrm{SNR}_c^{(\gamma=1.0)} = \mathrm{SNR}_0 \cdot G_p \text{ with } G_p = \int_{-\pi}^{\pi} \left|\frac{1}{1-A(\Omega)}\right|^2 d\Omega.$$

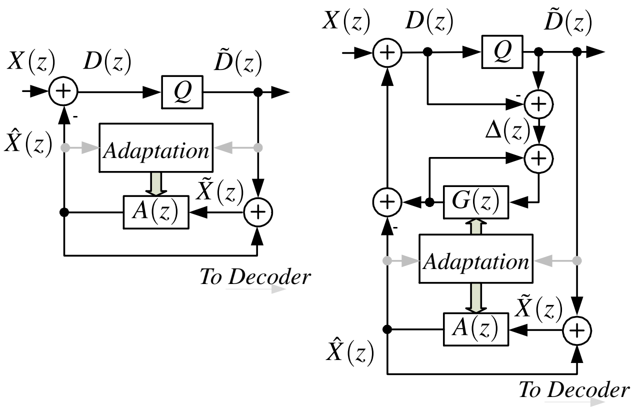
For any other choice of $0 \le \gamma < 1$, the overall SNR will be lower but the spectral shape of the quantization error within the reconstructed signal will be more and more alike the spectral envelope of the signal to be quantized. The latter leads to benefits in terms of the perceived quality taking into account the masking properties of the human auditory system. The other extreme case would be to set $\gamma = 0$. In that case, the quantization SNR is

$$\mathrm{SNR}_c^{(\gamma=0.0)} = \mathrm{SNR}_0 \tag{6}$$

which means that no benefit in terms of SNR on the one hand but the maximum noise shaping impact on the other hand can be achieved.

2.2 LPC with Backward Adaptation (B-LPC)

In LPC with backward adaptation, instead of finding the optimal solution given a short segment of the input signal, the LP coefficients are updated following a gradient descent algorithm on a sample-by-sample basis. The corresponding block diagram is shown in Figure 3 a). Mainly two methods have been described in the literature, the *Normalized Least Mean Square* (NLMS) algorithm and the *Gradient Adaptive Lattice* (GAL) algorithm. Both approaches are based on the computation of an instantaneous gradient which is then used to update the involved prediction coefficients such that the variance of the LP residual signal $d(k)$ is minimized.



a) Conventional Approach b) Novel Approach

Figure 3: LPC encoder with backward adaptation, conventional approach without noise shaping (a)) and novel approach with noise shaping (b)).

In analogy to LPC with forward adaptation, signal $\hat{X}(z)$ is the output from the predictor and an estimate for the input signal $X(z)$, and the LP residual signal is computed as $D(z) = X(z) - \hat{X}(z)$. However, $\hat{X}(z)$ is derived from the *previously quantized and reconstructed* input signal in case of backward adaptation as $\hat{X}(z) = A(z) \cdot \tilde{X}(z)$. $\tilde{X}(z)$ is available in the encoder as well as in the decoder so that no LP coefficients need to be transmitted in addition to the quantized signal $\tilde{D}(z)$ to reconstruct the input.
The sequential adaptation rules for both, the NLMS algorithm and the GAL algorithm, are described in [9], [10] and [11], [12], respectively, and shall not be discussed in detail here. The main difference is that in the NLMS approach, the predictor is specified as an FIR filter of order N_{LPC} whereas in the GAL approach, a lattice structure is employed to realize the predictor filter $A(z)$ containing N_{LPC} lattice stages. However, the lattice structure can be transformed into an FIR filter, the main difference therefore lies in different adaptation characteristics of the two proposed methods. Since no side information is required to be transmitted, predictors even of very high order can be realized without increasing the bit rate.

3 Backward Adaptive LPC with Noise Shaping (B-LPC-NS)

The novel LPC encoder with backward adaptation and noise shaping is shown in Figure 3 b). The main difference lies in the additional inner path to process the quantization error $\Delta(z)$ by means of filter $G(z)$. In this case, the encoder output can be derived as

$$\tilde{D}(z) = X(z) \cdot [1 - A(z)] + \Delta(z) \cdot \left[\frac{1 - A(z)}{1 - G(z)}\right] \qquad (7)$$

Due to the reconstruction in the decoder involving the LP synthesis filter, the decoder output signal is

$$\tilde{X}(z) = X(z) + \Delta(z) \cdot \left[\frac{1}{1 - G(z)}\right] \qquad (8)$$

Based on this result, the power spectral density of the quantization noise can be expressed as

$$\phi_{nn}^{(\text{B-LPC-NS})}(\Omega) = \phi_{\delta\delta}(\Omega) \cdot \left|\frac{1}{1 - G(\Omega)}\right|^2. \qquad (9)$$

Again, the noise shaping filter can be derived from $A(z)$ according to the rule as given in (3) as $G(z) = A(z/\kappa)$ with the noise shaping factor κ.
Even though the principle of noise shaping in (9) is almost identical to that in (5), the parametrization of the noise-shaping in B-LPC-NS deviates from that in F-LPC: This difference can be easily understood by reviewing the impact of $\gamma = \kappa = 1$: In that case, in F-LPC, the overall coding performance benefits in terms of the SNR but leads to a spectrally flat quantization noise. In contrast to this, in B-LPC-NS, the opposite happens for $\kappa = 1$: No benefit in terms of the SNR is achieved but the quantization noise within the reconstructed input signal is spectrally shaped in a way very similar to the signal to be quantized. Note also that approach B-LPC from Section 2.2 is a special case of B-LPC-NS for $\kappa = 0$.

4 Evaluations

For the evaluation of the novel approach, a simple codec was realized based on approach B-LPC-NS. The F-LPC approach is not considered in the following as a fair comparison would also involve the quantization of the LP coefficients which has not been realized in our simple test codec. Two different LP orders were investigated, $N_{\text{LPC}} = 9$ and $N_{\text{LPC}} = 18$. The quantizer involved in our test codec is based on the principle commonly known as Adaptive Quantization Backward (AQB) with a backward adaptive stepsize computation, normalization, and uniform scalar quantization of the normalized signal as specified in [3]. Noise shaping filter $G(z)$ is derived from $A(z)$ on a sample-by-sample basis.
At first, values of SNR_0 were measured for 5, 4 and 3 bits/sample to characterize the quantizer:

- $\text{SNR}_0 \approx 26$ dB for 5 bits/sample
- $\text{SNR}_0 \approx 20$ dB for 4 bits/sample
- $\text{SNR}_0 \approx 14$ dB for 3 bits/sample

For our evaluations, an artificial input signal was generated based on an auto-regressive (AR) model involving a fixed set of filter coefficients. Assuming that in the best case, the LP analysis filter identifies the given exemplary AR model filter coefficients, the maximum prediction gain was computed to be $10\log_{10}(G_p) \approx 13$ dB. The spectral envelope of the input signal as well as the spectral envelope of the quantization noise within the reconstructed signal were determined by finding approximations of the power spectral densities by averaging periodograms computed for overlapping segments of the respective signals for different values of κ. It turned out that GAL and NLMS approaches principally behave in a similar way and that the results for all bit rates and predictor orders are qualitatively very similar. Therefore, the resulting curves shown in Figure 4 are only for the GAL algorithm, a bit rate of 3 bits per sample and $N_{\text{LPC}} = 18$. In Figure 5, the measured values of SNR_c are shown as a function of the noise shaping parameter κ. Both figures confirm that the new approach for noise shaping has the desired impact: The parameter κ controls the spectral envelope of the quantization error within the reconstructed output signal as well as the SNR benefit as defined in (2). For higher values of κ, the spectral envelope of the quantization error is similar to that of the input signal (AR model) but the SNR is low whereas for lower values, the SNR is high but the quantization error is spectrally flat. Next, our simple test codec was used to process signals taken from a database of speech files recorded at a sample

ISBN 978-3-8007-3640-9

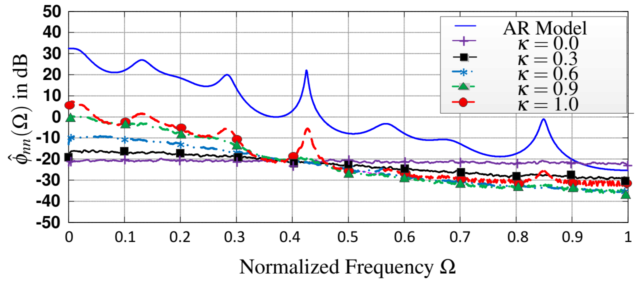


Figure 4: Spectral envelope defined by the AR model and measured approximations of power spectral density $\phi_{nn}^{(\kappa)}(\Omega)$ for different noise shaping factors κ, 3 bits/sample, GAL approach, $N_{\text{LPC}} = 18$.

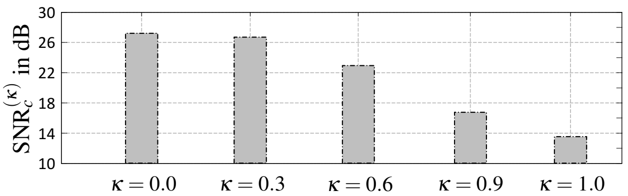


Figure 5: Measured values $\text{SNR}_c^{(\kappa)}$ for different noise shaping factors κ, 3 bits per sample, GAL approach.

rate of $f_s = 8$ kHz. In order to assess the perceived quality of the reconstructed signal, the processed speech files were rated based on the *Perceptual Evaluation of Speech Quality* (PESQ) measure [13], a measure to model subjective listening tests. Informal listening tests conducted prior to the assessment showed that a value of $\kappa = 0.7$ seems to be the best compromise to achieve a high coding SNR as well as a good spectral shaping of the quantization error with respect to human perception.

The resulting average PESQ values for the conventional approach without noise shaping (B-LPC identical to B-LPC-NS with $\kappa = 0.0$) and for the new approach (B-LPC-NS with $\kappa = 0.7$) were compared to derive Δ_{PESQ} as a measure of the **quality improvement** due to the involved noise shaping. The values are shown in Figure 6 for the three bit rates of 24, 32, and 40 kBits/s for the GAL as well as for the NLMS based adaption rule. The corresponding absolute PESQ values for the GAL approach for the investigated bit rate is shown below the bit rates. In this test, the GAL approach principally showed a higher performance than the NLMS approach. The new approach (B-LPC-NS) outperforms the conventional approach (B-LPC) consistently with a significant maximum increase of approximately 0.2 on the PESQ scale.

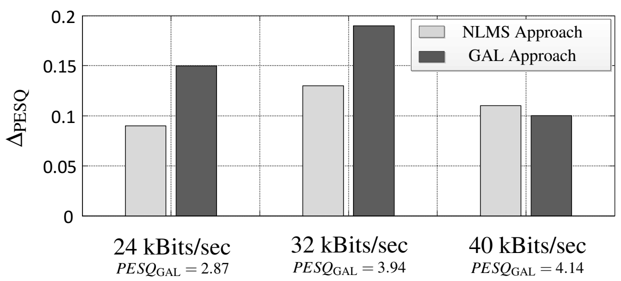


Figure 6: Measured increase of PESQ values due to noise shaping, $\kappa = 0.7$.

5 Conclusions

LPC with forward block adaptation of the LP coefficients requires that the LP coefficients are transmitted as side information but allows for noise shaping to better account for human perception. LPC with backward sequential adaptation of the LP coefficients does not require any transmission of LP coefficients and therefore saves bit rate. However, the lack of noise shaping has been been a drawback in terms of the achievable perceived quality in the past.

In order to overcome this shortcoming, in this contribution, a novel approach for noise shaping in LPC with backward adaptation was proposed. It was shown that the achieved noise shaping effect principally is very similar to that known from LPC with forward adaptation with the exception that the parametrization is somewhat different. In the end it turns out in evaluations based on a simple test codec that an increase in perceived quality can be achieved by the novel approach compared to the standard realizations of LPC with backward adaptation: At an overall data rate of 32 kBits/s for speech signals recorded with a sample rate of $f_s = 8$ kHz, the new approach outperforms the conventional approach by a difference of almost 0.2 in PESQ.

References

[1] ETSI, Rec. GSM 06.90, "Digital Cellular Telecommunications System (Phase 2+); Adaptive Multi-Rate (AMR) Speech Transcoding." version 7.2.1, rel. 1998, Apr. 2000.

[2] ETSI/3GPP, Rec. GSM 26.190/ITU-T G.722.2, "Adaptive Multi-Rate Wideband Speech Transc. (AMR-WB)," 2001.

[3] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Prentice Hall, 1984.

[4] M. R. Schroeder, B. S. Atal, and J. L. Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear," *Acoustical Society of America*, pp. 1647–1652, Dec. 1979.

[5] ITU, *Security architecture for Open Systems Interconnection for CCITT applications (ITU-T Recommendation G.726)*. International Telecomm. Union, Mar. 1991.

[6] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding and Error Concealment*. John Wiley & Sons, 2006.

[7] W. B. Kleijn and K. K. Paliwal, *Speech Coding and Synthesis*. Amsterdam: Elsevier, 1995.

[8] H. Krüger, B. Geiser, and P. Vary, "On noise propogation in closed-loop linear predictive coding," *Accepted for publication at the ICASSP 2014*, 2014.

[9] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. J. Jr, "Stationary and nonstationary learning characteristics of lms adaptive filter," *Proceedings of IEEE*, vol. 64, pp. 1151–1162, Aug. 1976.

[10] S. Haykin, *Adaptive Filter Theory*. Prentice Hall, 3 ed., 1996.

[11] B. Friedlander, "Lattice filters for adaptive processing," *Proceedings of IEEE*, vol. 70, pp. 829–867, Aug. 1982.

[12] L. J. Griffiths, "An adaptive lattice structure for noise-cancelling applications," *IEEE Trans. on Ac., Speeech and Signal Processing*, vol. 3, pp. 87–90, Apr. 1978.

[13] ITU-T Rec. P.862, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," 2001.

