EVALUATION OF A WARPED LINEAR PREDICTIVE CODING SCHEME

Aki Härmä

Helsinki University of Technology
Laboratory of Acoustics and Audio Signal Processing
P. O. Box 3000, 02015, Espoo, Finland
Aki.Harma@hut.fi

ABSTRACT

Basically all conventional digital signal processing techniques can be warped by introducing a simple modification to the system. In this paper, the focus is in warped linear predictive coding techniques with application to speech and audio coding. The performance of warped LPC is compared with a conventional LPC in listening tests and in terms of technical measures. This is done at various sampling rates as a function of the order of the LPC model.

1. INTRODUCTION

Nonuniform resolution FFT was introduced by Oppenheim, Johnson, and Steiglitz [10]. The main idea was to use a network of cascaded first order allpass sections for frequency warping of the signal and then apply Fast Fourier Transform (FFT) to produce the warped spectrum from the preprocessed signal.

The transfer function of a first order allpass, AP, filter is given by

$$D(z) = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}}, \tag{1}$$

By definition, the magnitude response of the filter is a constant. The phase response of $D(z)$ is given by

$$\tilde{\omega} = \omega + 2\arctan\left(\frac{\lambda \sin(\omega)}{1 - \lambda\cos(\omega)}\right). \tag{2}$$

The phase function determines a frequency mapping occurring in the allpass chain [10, 14]. For a certain value of λ the frequency transformation closely resembles the frequency mapping occurring in the human auditory system. Smith and Abel [12, 13] derived an analytic expression for λ so that the mapping, for a given sampling frequency f_s, matches the psychoacoustic *Bark*-scale mapping. The value is given by

$$\lambda_{f_s} \approx 1.0211(\frac{2}{\pi}\arctan(0.076 f_s))^{1/2} - 0.19877. \tag{3}$$

The value of λ obeys this formula in all experiments reported in this article.

2. WARPED LINEAR PREDICTIVE CODING

In classical forward linear prediction [8] an estimate for the next sample value $x(n)$ is obtained as a linear combination of N previous values given by

$$\hat{x}(n) = \sum_{k=1}^{N} a_k x(n-k), \text{ or } \hat{X}(z) = [\sum_{k=1}^{N} a_k z^{-k}]X(z), \tag{4}$$

where a_k are fixed filter coefficients. Here z^{-1} is a simple *unit delay filter* or a *shift operator*, which may be generalized using a first-order allpass, AP, filter $D(z)$ to obtain

$$\hat{X}(z) = [\sum_{k=1}^{N} a_k D(z)^k]X(z). \tag{5}$$

In the time domain, $D(z)^k$ can be interpreted as a *Generalized Shift Operator* defined as

$$d_k[x(n)] \equiv \underbrace{h(n) * h(n) * \cdots * h(n)}_{k-\text{fold convolution}} * x(n), \tag{6}$$

where the asterisk is a convolution and $h(n)$ is the impulse response of $D(z)$. Furthermore, we denote $d_0[x(n)] \equiv x(n)$. The minimum mean square error of the estimate may now be written as

$$e = E\left[|x(n) - \sum_{k=1}^{N} a_k d_k[x(n)]|^2\right], \tag{7}$$

where $E[\cdot]$ is expectation. Minimization of this with $\partial e/\partial a_k = 0$ and $k = 1, 2, \cdots, N$ leads to a system of *normal equations*

$$E[d_j[x(n)]d_0[x(n)]] - \sum_{k=1}^{N} a_k E[d_k[x(n)]d_j[x(n)]] = 0, \tag{8}$$

with $j = 0, \cdots, N-1$. Since $H(z)$ is a linear filter, it is straightforward to show that

$$E[d_j[x(n)]d_k[x(n)]] = E[d_{j+p}[x(n)]d_{k+p}[x(n)]], \forall j, k, p, \tag{9}$$

which means that the same correlation values appear in the both parts of (8). Therefore (8) can be seen as a generalized form of *Wiener-Hopf* equations. The correlation terms can be easily computed and *optimal* coefficients a_k can be solved efficiently using, e.g, *Levinson-Durbin* algorithm equally as in the conventional autocorrelation method of linear prediction. Correspondingly, we now have a *prediction error filter* given by

$$A(z) = 1 - \sum_{k=1}^{N} a_k D(z)^k, \tag{10}$$

which can be implemented directly by replacing all the unit delays of a conventional FIR structure with $D(z)$ blocks. It is also possible to implement a *synthesis filter* given by

$$A^{-1}(z) = \frac{1}{1 - \sum_{k=1}^{N} a_k D(z)^k}, \tag{11}$$

using, e.g., techniques presented in [1, 2].

Strube [14] pointed out that in the frequency domain the prediction error power of (7) takes the following form:

$$\sigma^2 = \int_{-\pi}^{\pi} E(\tilde{\omega}) d\tilde{\omega} = \int_{-\pi}^{\pi} E(\omega) \frac{1-\lambda^2}{1-\lambda^2-2a\cos(\omega)} d\omega \quad (12)$$

This can be derived from (7) using Parceval's theorem and (2), and its derivative $d\tilde{\omega}/d\omega$.

This means that (8) minimizes an error weighted with $(1-\lambda^2)/(1-\lambda^2-2a\cos(\omega))$. The transfer function of this weighting filter is given by

$$D_0(z) = \frac{\sqrt{1-\lambda^2}}{1-\lambda z^{-1}}. \quad (13)$$

This is a first-order lowpass filter. In practice this causes that the spectrum at the output of (10) is not perfectly flat but it has lowpass characteristics. In all experiments in this article, the residual signal $r(n) = x(n) - \hat{x}(n)$ is filtered using $D_0^{-1}(z)$ to produce a flat residual spectrum for the quantizer. This is a stable filter because $0 < \lambda < 1$. Moreover, $D_0(z)$ is applied for the excitation before synthesis filtering. This is done in order to make the comparison between prediction gain and spectral flatness measures reasonable in Section 4.

3. TEST SETUP

3.1. Simulated codec

In this paper, the performance of warped linear predictive coding is compared with conventional linear predictive coding. This is done in a simulated *residual-driven* codec where the autocorrelation method of linear prediction is used to estimate the coefficients of the warped filter and quantization process is simulated by adding white noise to the excitation signal in the synthesis phase. It is assumed here that results with this simplified LPC scheme reflect also results that that could be obtained by comparing any modern LPC based speech or audio codec, e.g., a CELP codec, and its warped version. As was pointed out earlier, almost any DSP algorithm can be warped.

The simulated encoder and decoder are shown in Figs. 1a and b, respectively. The computation of coefficients is performed in frames of 20 ms. In the encoder, the coefficients are used in a prediction error filter to produce a residual signal. This signal is quantized using Jayant's one-word memory quantizer [3]. In this simulated setup, the role of the quantizer is used to produce a noise signal which is obtained by subtracting the original residual from quantized residual, as shown in Fig. 1b. After the synthesis filtering the quantization noise has a spectral shape determined by the synthesis filter as usual in D*PCM codecs.

The coefficients of the filter are computed in frames of 20 ms using a Hamming window. The analysis is overlapping such that an analysis frame starts after every 10 ms interval. The coefficients of the filter are not quantized and no bandwidth expansion or other techniques are applied to the obtained all-pole model. Filter coefficients are expressed as reflection coefficients of a corresponding warped lattice filter and they are linearly interpolated between adjacent frames using a trapezoidal rule. Filters are implemented in the warped lattice form [1, 2].

In the decoder, see Fig. 1b, the quantized residual is first subtracted from the original residual to produce a quantization error signal $q(n) = r(n) - \hat{r}(n)$, which is approximately white noise but follows roughly the energy envelope of the original signal. The excitation for the time-varying synthesis filter (11) is a weighted sum of the original residual and the quantization error signal given by

$$\tilde{r}(n) = r(n) + \sqrt{10^{-SNR/10}} G_r q(n), \quad (14)$$

where G_r is a gain coefficient which is used to scale $q(n)$ so that $E[|G_r q(n)|^2] = E[|r(n)|^2]$. In listening tests, a subject may adjust the parameter SNR in real time to find the threshold of audibility for the quantization noise in the presence of the signal. The parameter SNR is the *Signal-to-Noise Ratio* for the residual signal and therefore it has, roughly, the following relation to the bit-rate of the quantizer:

$$SNR/\mathrm{dB} = 6b + \delta, \quad (15)$$

where b is the number of bits and δ is some constant, see, e.g., [11].

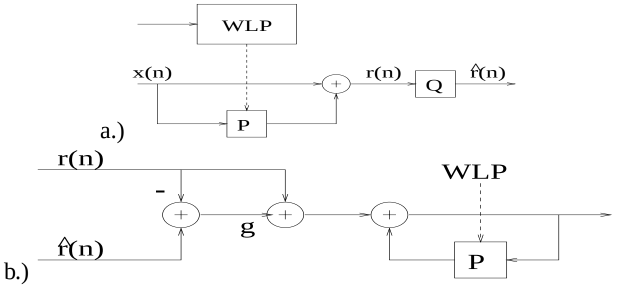

Figure 1: a) Simulated encoder. b) Simulated decoder

id	Sequence	identifier	source
1	English horn	E4 note	MUMS v1 09-13
2	Tubular bells	G4 note	MUMS v2 10-08
3	Electric guitar	chord	MUMS v8 04-02
4	Triangle	onset removed	MUMS v3 12-25
5	Violin	D4 open	MUMS v1 01-09
6	Female voice	singing /a/	Lab. of Acoustics
7	Male voice	/oe/	Lab of Acoustics
8	Trumpet	a long note	MUMS v2 16-12

Table 1: Test sequences. MUMS refers to the McGill University Master Sample collection.

3.2. Test sequences

The choice of test sequences plays an important role in designing the test setup. The listening test results were collected using steady-state segments from 8 music and voice signals. The test signals are listed in Table 1. Most of the test sequences are anechoic recordings from the McGill University Master Sample [9] collection. The signals were chosen so that they represent a wide range of clearly identifiable musical or speech sounds. Another criteria at this phase was to find sequences for which the variance in listening tests among different listeners and between trials is small. Some traditional test sequences, e.g., the harpsichord and some other *noisy* sounds, were left out of the set because it turned out that it was difficult for the listeners to judge the quality accurately, i.e., the variance was high.

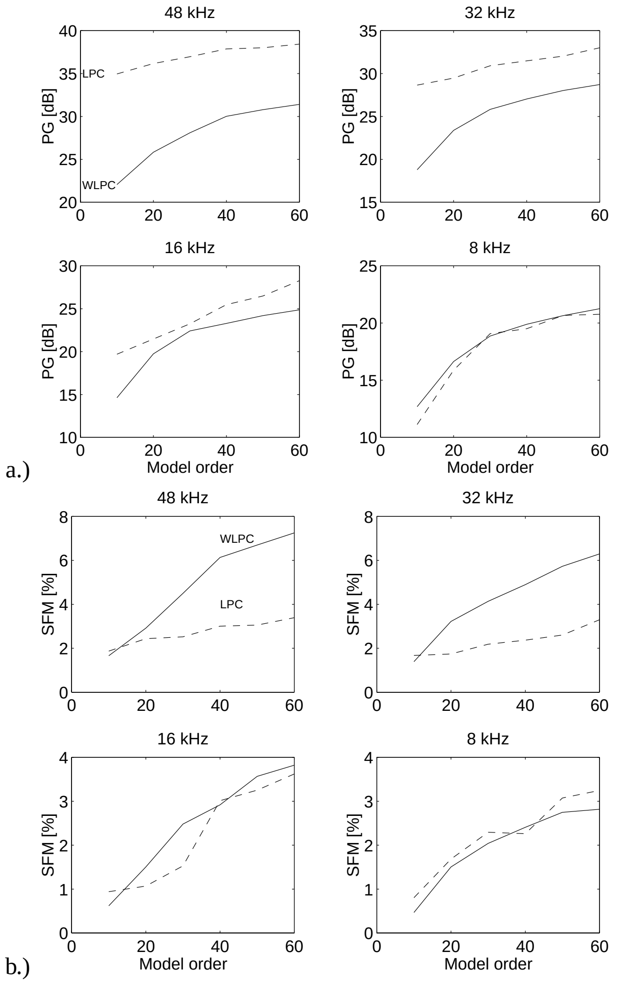

Figure 2: a) Prediction gain, and b) Spectral flatness values in a warped LPC (solid curve) and a conventional LPC (dashed curve) as a function of LPC order at four different sampling rates.

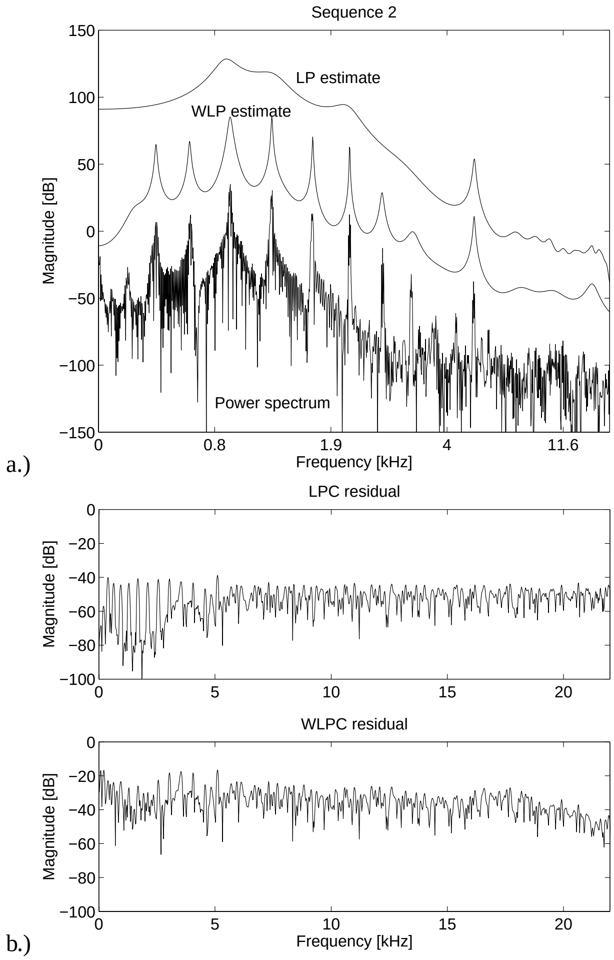

Figure 3: a) Power spectrum for a 1024 excerpt from test sequence 2 at 48 kHz sampling rate. 40th order LP and WLP spectral estimates are plotted in the same figure. The frequency scale is warped. b.) Residual spectra in LPC and WLPC.

4. PERFORMANCE IN TERMS OF TECHNICAL MEASURES

The minimization of the mean square error of the prediction error signal is equivalent with maximization of *prediction gain* measure [4]. Another widely used measure for the performance of an LPC model is based on measuring the whitening property of an inverse filter. Spectral flatness measure, SFM, is usually expressed as a ratio between geometric and arithmetic averages of a power spectrum of the residual signal.

The average prediction gain over the test sequences in Table 1 in a warped (solid curve) and a conventional LPC (dashed curve) as a function of model order at four different sampling rates are shown in Figs. 2a. The corresponding SFM percentages are shown in Fig. 2b. The prediction gain in WLPC is lower than in LPC at 48, 32, and 16 kHz sampling rates. However, the spectral flatness measure gives clearly higher values for WLPC than for LPC at those sampling rates. At 8 kHz sampling rate the values for both measures almost coincide. The difference between prediction gains in WLPC and LPC decreases as the model order increases. At 48 and 32 kHz sampling rates the spectral flatness increases faster than in LPC as a function of model order. Therefore, it the SFM favours WLPC especially if the model order is high.

Figure 3a shows the power spectrum of a 1024 sample (at the sampling rate of 48 kHz) excerpt of Test Sequence 2, *Tubular Bell* and the estimated 40th order LPC and WLPC spectra. The frequency axis is warped so that it approximates the Bark scale. The warped model can pick most of the peaks at low frequencies while the conventional model is probably too accurate at high frequencies. Fig. 3b shows the corresponding residual spectra in the two cases. Both inverse filters reduce spectral level at low frequencies approximately by the same amount but WLP model removes the spectral peaks while LPC model, as one can see by Fig. 3a, only whitens the signal in a coarse sense, i.e., the peaks of the original spectrum are almost unchanged. This is one reason why LP gives a higher prediction gain even if the model in the spectral sense is worse than in WLP. Another thing is that the overall level of the residual spectrum in WLPC is at a higher level.

One can see from Fig. 3a that the spectral resolution of WLP is higher than in LP below approximately 5 kHz and lower above that.

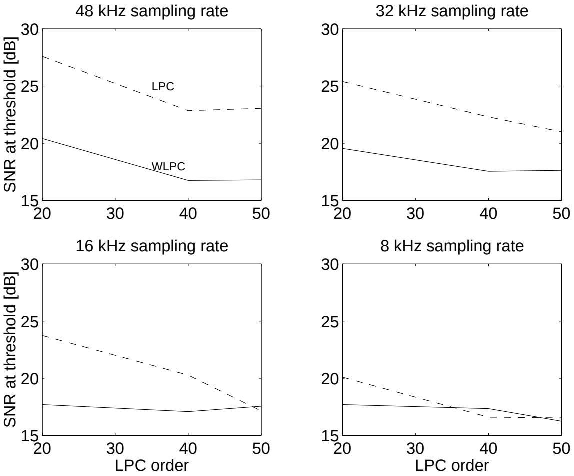

Figure 4: Average listening test results

5. LISTENING TESTS

The preliminary listening test reported in this paper was done in a standard listening room [5] using one Genelec 1032 loudspeaker. Test sounds were played by a Silicon Graphics Indigo workstation. The two listeners used a computer mouse to adjust a slider, corresponding to the SNR parameter, in a graphical user interface. Basically, the test procedure is a version of *the method of adjustment*. In this real-time system a listener may freely adjust the SNR parameter and hear the difference immediately. The goal was to find the threshold of audibility for quantization noise.

The test material consisted of the eight steady-state musical and speech sounds listed in Table 1. The duration of each sample was one second but the sounds were played in a continuous loop. The warped and conventional LPC simulations were tested at four different sampling rates (8, 16, 32, and 48 kHz) and with three different orders of the LPC or WLPC model, i.e, 20, 40, and 50. In this preliminary test, only one subject was tested in all 12 experiments. Two other listeners participated in some of these listening tests.

The average listening test results over all test samples are shown in Fig. 4. At sampling rates of 48 kHz and 32 kHz, the SNR for residual in the warped LPC is approximately 6 dB below that of a conventional LPC. According to (15) this means that a sufficient bitrate for residual in WLPC is one bit per sample less, i.e., 48kb/s or 32 kb/s less, than in LPC. At the 16 and 8 kHz sampling rates the difference between WLPC and LPC is a decreasing function of model order. In the case of 50th order model at the 16 kHz sampling rate, or 35th order model at the 8 kHz sampling rate the use of warped LPC brings no gain compared to the conventional case. However, below that the difference is clear.

It is not shown in the figures, but the curves can be almost linearly extrapolated towards lower orders of the model. For example, at 8 kHz sampling rate the difference between a 10th order conventional and a warped LPC model is around 5 dB. Basically, this means that the order of the model can be significantly lower in the warped LPC compared to the conventional LPC. Similar results with narrow band speech coding were also obtained by Krüger and Strube [7] and, e.g., Koishida et al. [6].

6. CONCLUSIONS AND FUTURE WORK

The purpose of this paper was to compare frequency-warped LPC and conventional LPC in terms of listening tests and technical measures as a function of model order and sampling rate. Preliminary listening test results seem to indicate that it is beneficial to use warped LPC especially if the sampling rate is above 8 kHz or if the order of the model is low.

At the time of writing this, the final listening tests are running and it can be expected that more accurate data can be presented in the conference.

7. ACKNOWLEDGMENT

This work has been supported by the Academy of Finland. The author is grateful to the listeners.

8. REFERENCES

[1] A Härmä. Implementation of recursive filters having delay free loops. In *Proc of ICASSP'98*, volume III, pages 1261–1264, Seattle, 1998.

[2] A. Härmä. Implementation of frequency-warped recursive filters. *Signal Processing (to appear)*, 80(3), 2000.

[3] N S Jayant. Adaptive quantization with one-word memory. *Bell Syst. Tech. J.*, pages 1119–1144, 1973.

[4] N. S. Jayant and P. Noll. *Digital coding of waveforms*. Prentige-Hall, New Jersey, 1984.

[5] A. Järvinen, L. Savioja, H. Möller, V. Ikonen, and A. Ruusuvuori. Design of a reference listening room—a case study. In *AES 103rd Convention preprint no. 4559*, New York, USA, September 1997. AES.

[6] K Koishida, K Tokuda, T Kobayashi, and S Imai. Celp coding system based on mel-generalized cepstral analysis. In *Proc. of ICSLP'96*, volume 1, 1996.

[7] E. Krüger and H. W. Strube. Linear prediction on a warped frequency scale. *IEEE Trans. Acoust. Speech, and Signal Proc.*, 36(9):1529–1531, September 1988.

[8] J. D. Markel and A. H. Gray. *Linear Prediction of Speech*, volume 12 of *Communication and Cybernetics*. Springer-Verlag, New York, 1976.

[9] F Opolko and J Wapnick. *McGill University Master Samples User's Manual*. McGill University Faculty of Music, Montreal, 1989.

[10] A V Oppenheim, D H Johnson, and K Steiglitz. Computation of spectra with unequal resolution using the fast Fourier transform. *Proc. of IEEE*, 59:299–301, 1971.

[11] L. R. Rabiner and R. W. Schafer. *Digital Processing of Speech Signals*. Prentice-Hall Inc., New Jersey, 1978.

[12] J O Smith and J S Abel. The Bark bilinear transform. In *Proc. of IEEE WASPAA*, New Paltz, 1995.

[13] J. O. Smith and J. S. Abel. Bark and ERB bilinear transform. *Trans. Speech and Audio Processing*, 7(6):697–708, November 1999.

[14] H W Strube. Linear prediction on a warped frequency scale. *J. of the Acoust. Soc. Am.*, 68(4):1071–1076, 1980.