

# Auditory Interpretation and Application of Warped Linear Prediction

Matti Karjalainen

Helsinki University of Technology  
Laboratory of Acoustics and Audio Signal Processing  
P.O. Box 3000, FIN-02015 HUT, Espoo, Finland  
matti.karjalainen@hut.fi

## Abstract

This paper presents a study on a specific signal processing technique, the warped linear prediction (WLP), interpreted here from an auditory perception point of view. WLP is a signal modeling technique that can approximate the frequency resolution properties of the human ear. WLP is shown in this paper to have system-level similarities with signal analysis found in the peripheral auditory system. Several phenomena in auditory modeling are addressed where WLP-based spectral modeling and residual processing are useful. Pitch analysis, including multi-pitch analysis, segregation of harmonic signals by pitch-predictive means, and other processing techniques (synthesis, recognition) are shortly discussed.

## 1. Introduction to WLP

Frequency-warped linear prediction (WLP), first introduced systematically by Strube [1], exhibits some interesting properties from the viewpoint of auditory modeling and related applications. In this contribution we discuss different aspects of WLP and its applications.

Warped linear prediction is a signal analysis and modeling technique, similar to ordinary linear prediction [2], but in a manner where the frequency resolution of spectral modeling is controllable by a single warping parameter [3]. The idea of a warped frequency scale and related resolution is based on using allpass sections instead of unit delays in DSP structures, i.e.,

$$\tilde{z}^{-1} = D_1(z) = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}} \quad (1)$$

where  $\lambda$ ,  $-1 < \lambda < 1$ , is a warping parameter and  $D_1(z)$  is a warped (dispersive) delay element. With a proper value of  $\lambda$  the warped frequency scale shows a good match to the psychoacoustically defined Bark scale, thus optimizing the frequency resolution from the point of view of auditory perception [4]. For example with a sampling rate of 22 kHz, Bark warping is obtained using  $\lambda \approx 0.63$ . WLP analysis is easily realized by modifying only the autocorrelation computation using a version where unit delays are replaced by allpass sections. The same holds for inverse filtering to obtain the residual (excitation) signal. The synthesis filter  $A(z)$  is

$$A(z) = \frac{1}{1 + \sum_{i=1}^N a_i D_1(z)^i} \quad (2)$$

where  $a_i$  are feedback coefficients of the recursive structure. This cannot, however, be realized in a straightforward manner since in recursive structures the replacement of Eq. (1) results in delay-free loops. Techniques to avoid this problem are

discussed for example in [5] and [6]. Otherwise, in most aspects WLP and frequency-warped signal processing are easily derived from traditional techniques<sup>1</sup>.

## 2. Basic properties of WLP

Among interesting properties of WLP are the ability of spectral modeling similar to loudness density spectrum (auditory spectrum) estimation, and the fact that information in the (inverse-filtered) WLP residual resembles the overall information in the auditory nerve firing. The latter fact is achieved by the WLP filter coefficients and a gain factor so that the residual shows spectral flattening and level compression similar to the adaptation of firing rate in the auditory nerve. Depending on the type of processing desired, a WLP front end for residual computation may be followed by a filterbank to separate the signal into critical bands. Often such separation is not needed at all in further processing.

The following figures illustrate the basic properties of WLP when applied to speech signals. Figure 1 shows a comparison of ordinary and warped linear prediction for different filter orders as applied to a vowel (Finnish /a/). While LP exhibits frequency-independent inherent spectral resolution, Bark-scaled WLP focuses best resolution on frequencies below 500 Hz and decreasing resolution above it. While LP order of about 24 is needed for sample rate of 22 kHz (and 46 for 44 kHz) in order to model average formant frequency density of 1 per kHz, much lower WLP orders are enough for the same resolution at lowest signal frequencies (about 14 for sampling rate of 22 kHz). Then the higher formants are not fully resolved anymore, corresponding to the auditory resolution principle<sup>2</sup>. Although there is no explicit auditory model involved in WLP, the obtained auditory spectrum exhibits relatively accurate excitation spreading and spectral masking properties, as demonstrated in [3] for WLP-based audio coding.

Figure 2 plots the spectra of vowel /a/ and fricative /s/ on the (normalized) Bark scale of Bark-WLP. Now the inherent resolution is uniform on the Bark scale. Thus this representation can be considered as an auditory spectrum, especially if the transfer functions of the external and middle ear were included before WLP analysis and the dB scale were encoded to loudness (sone) scale.

The WLP residual (input signal inverse-filtered by WLP predictor filter) for vowel /a/ is depicted in Fig. 3b along with the original signal waveform Fig. 3a. The residual looks much like an ordinary LP residual. It is sharply peaked at the glottal

<sup>1</sup>A Matlab toolbox for warped signal processing is available in <http://www.acoustics.hut.fi/software/warp>.

<sup>2</sup>The overall spectral resolution depends on the WLP order and windowing used for autocorrelation computation.

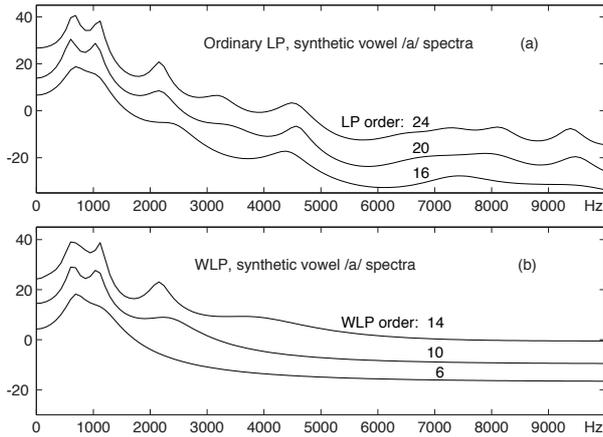


Figure 1: (a) LP and (b) WLP spectra of vowel /a/ for different filter orders.

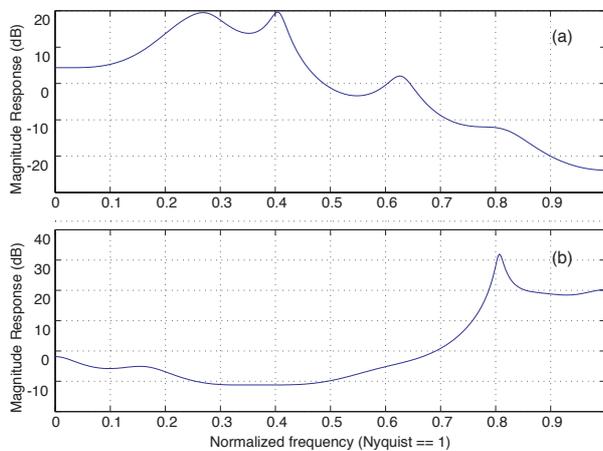


Figure 2: WLP spectra of (a) vowel /a/ and (b) fricative /s/ on a normalized Bark scale.

closure time and shows ripples after that, but the oscillation due to strong low-frequency formants of the original waveform is clearly reduced.

Spectral properties of the WLP residual are illustrated in Fig. 4b. Spectral whitening is obvious (as in ordinary LP), although lowest valleys of spectrum are not fully flattened. This spectral flattening tendency resembles the adaptation process taking place in auditory nerve firing rates.

An interesting finding is that the residual, although spectrally flattened, still conveys speech-related information in addition to pitch and periodicity. Although sounding buzzy, it conveys features of individual speaker and to some degree also phone identity: the spoken message can in most cases be recognized even when listening to the residual only. From such experiments it may be assumed that the time-frequency analysis carried out in the auditory system is roughly equivalent to the combination of the spectral information of WLP spectrum and the temporal fine structure of the WLP residual.

The residual represents roughly the overall information carried by the auditory nerve fibers. It can be split into critical bands by a simple bandpass filterbank if needed in further analysis. In some applications it is enough to split the residual into two parts. Below about 1–2 kHz the neural firings in the au-

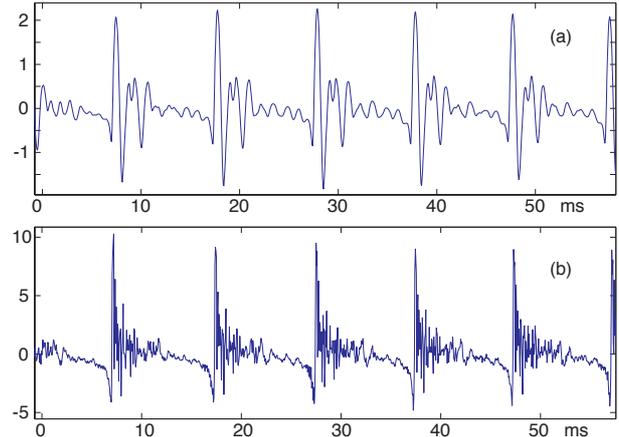


Figure 3: (a) WLP residual signal and (b) vowel /a/.

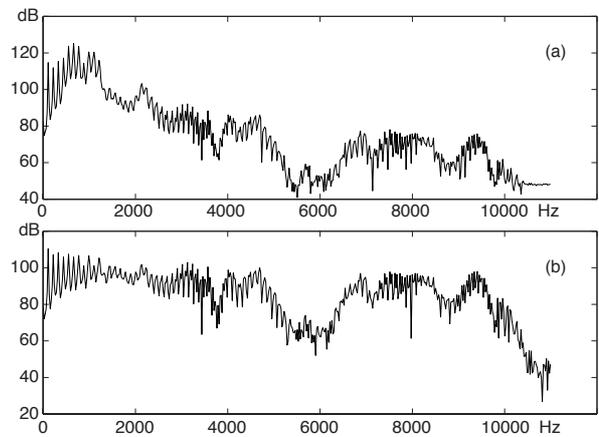


Figure 4: Spectral flattening of WLP residual: (a) spectrum of vowel /a/ and (b) WLP residual spectrum of the same vowel.

ditary nerve are (in statistical sense) synchronous to half-wave rectified and low-pass filtered critical band signals. Together this frequency range conveys information that is similar to the low-pass filtered WLP residual. Figure 5b illustrates this for the case of vowel /a/ with 1 kHz low-pass filtering. Figure 5a depicts the high-frequency part of WLP residual half-wave rectified and then low-pass filtered. This resembles the overall neural information of high-frequency neural channels that are not anymore fully synchronized to critical band waveform but rather to its temporal envelope.

### 3. Auditory spectrogram through WLP

Figure 6 illustrates the use of WLP in auditory spectrogram computation. WLP of order 24 has been computed in this case with a hop-size of 2 ms (and 25 ms window) for high presentation resolution. Notice that the linear prediction gain parameter is not taken into account so that a level normalization (0 dB average) in the spectrogram takes place.

In time-varying signals, particularly in onsets and transient sounds the auditory system exhibits two complementary forms of temporal processing: overshoot in neural firing rate prior to adaptation to steady-state level and temporal integration in loudness formation. In order to include this in a WLP analysis system, an adaptive version of WLP is needed. Notice that a

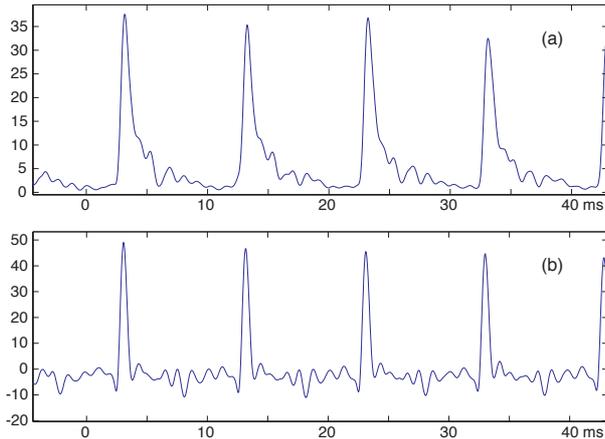


Figure 5: Residual from vowel /a/ split into two parts: (a) 1.0 kHz high-pass filtered, rectified, and then 1 kHz lowpass filtered (= envelope following); (b) 1 kHz low-pass filtered (= waveform following).

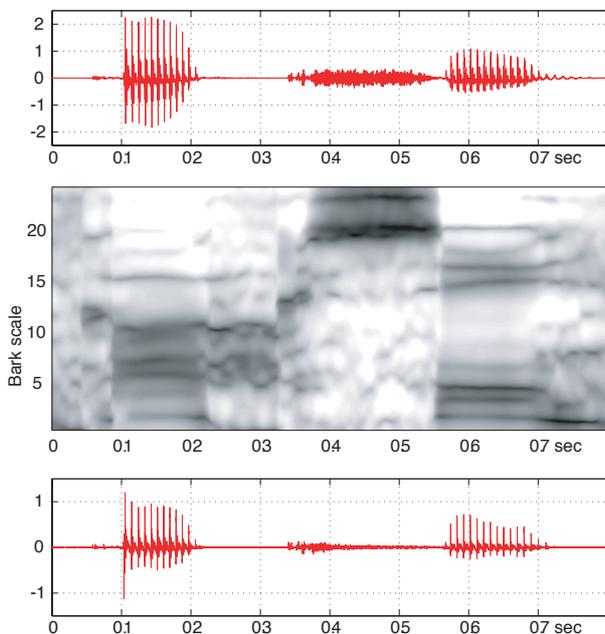


Figure 6: Word /kaksi/ (top), its auditory spectrogram by WLP (middle), and WLP residual (bottom).

straightforward adaptive WLP does not implement neither overshoot and temporal integration nor proper spectral changes. A feedforward control of WLP inverse filter by solving the filter coefficients from temporally smoothed warped power spectrum is a better strategy for this. The realization of such a property remains a future task.

#### 4. WLP support for source segregation

The ability to segregate sound sources is one of the advanced features of auditory perception that has been difficult to simulate by machines. We have applied WLP analysis to the task of segregating voiced speech signals, although this is only partially motivated by auditory principles.

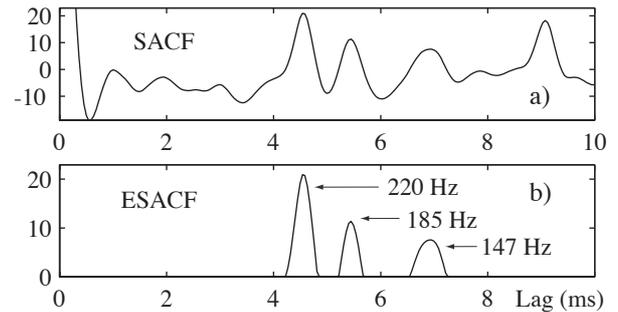


Figure 7: Multi-pitch analysis of a mixture of three harmonic sounds: (a) summary autocorrelation function (SACF), (b) enhanced summary autocorrelation (ESACF), see text.

The first part of the task is to do multi-pitch analysis, assuming that the signal is an additive mixture of harmonic sounds. In [7] we have proposed a multi-pitch analysis method that first applies WLP whitening of the input signal, and then does the following procedure. WLP residual is split into two parts as described above at the end of Section 2. A special version of autocorrelation is computed for both low-frequency and high-frequency parts and these are summed (SACF = summary autocorrelation). A technique to remove the periodic structure of autocorrelation leads to enhanced SACF (ESACF) function that is a useful periodicity lag representation for multi-pitch analysis. Figure 7 depicts an example where three harmonic signals with different fundamental frequencies are resolved by this technique.

In [8] we demonstrated a method to separate two vowels of different fundamental frequency. The goal was to estimate the individual spectra of the vowels. After estimating each one of two pitches of a mixture, pitch-predictive canceling of one excitation signal was applied to achieve the other one. This was repeated for the second second excitation in a similar way. Then the excitations were applied to the WLP synthesis filter and the resulting signals were again LPC-modeled in order to estimate the individual spectral envelopes.

Here we just demonstrate the ability of WLP residual to help in the segregation process. Figure 8a plots the full WLP residual of the mixture signal of Finnish vowels /a/ and /i/. When the WLP residual is split into low- and high-frequency components they look as in Figures 8b and c. Now the two sequences of excitation peaks are more clearly observable since the formant structure of the vowels does not disturb so much and the temporal fine structure of full residual is reduced, which helps in further processing, such as in different segregation techniques.

#### 5. Other WLP applications

The auditory properties of WLP processing have already been utilized in many speech and audio applications [3]. In speech processing the principle has been applied for example to speech synthesis [9] where the main advantage is due to reduced synthesis filter order in source-filter modeling of speech signals. This helps in generating the control parameters of the synthesis filter.

Another application is speech recognition where we have shown that WLP-type of preprocessing is a useful method [10]. We have found that WLP processing is a method competitive in efficiency and accuracy with mel-cepstral coefficients.

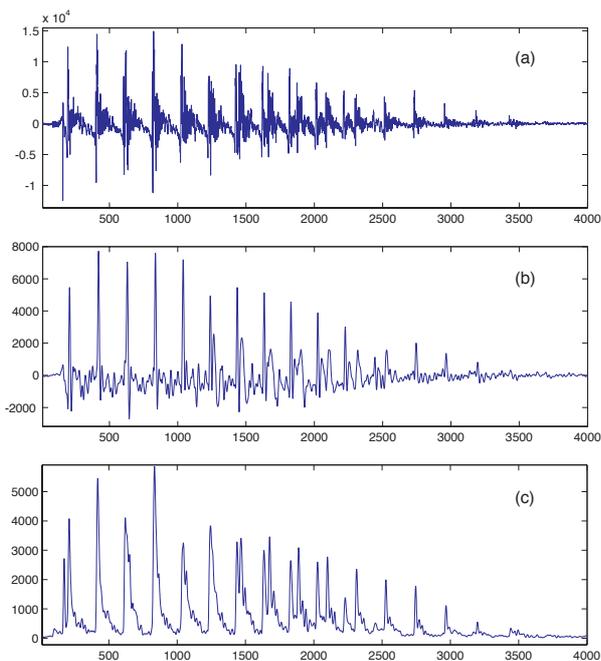


Figure 8: /a/ & /i/ mixed WLP residuals: (a) full residual, (b) low band residual, (c) high band residual.

Warped linear prediction can be used furthermore in speech coding [1] and particularly in audio applications [3].

## 6. Discussion and Conclusions

This paper has discussed the properties of warped linear prediction in relation to auditory signal processing. It is demonstrated that Bark-WLP is able to approximate some of the main auditory features, such as representing spectral information on the Bark scale and representing the overall auditory nerve firing rate information (or critical band firing rates if a filterbank is added) including spectral flattening in steady-state signals. Since WLP is a relatively straightforward DSP method resembling ordinary linear prediction, it is a potential technique where auditory principles have to be integrated in conventional DSP frameworks.

The examples given in this paper illustrate selected cases where the properties of WLP are prominent. In complex cases, especially in source segregation and auditory scene analysis applications, more advanced versions of the principles described here are needed. Many special cases have also to be worked out to find the capabilities and limitations of the technique. Such cases include for example: application of WLP to source separation and modeling for object-based audio coding (extensions to audio coding discussed in [3]), analysis and modeling of reverberant signals (separation of source and acoustic environment), robust segregation of voice from background noise for speech recognition, etc.

In conclusion, WLP is found to be a versatile DSP technique that combines traditional DSP and basic auditory properties in a flexible way.

## 7. Acknowledgments

This work was started while the author had a sabbatical academic year 1999–2000 financed by the Academy of Finland.

## 8. References

- [1] Strube H. W., "Linear Prediction on a Warped Frequency Scale," *J. Acoust. Soc. Am.*, vol. 68, no. 4 (1980).
- [2] Markel J. D., and Gray A. H., *Linear Prediction of Speech*, Springer Verlag, New York, 1976.
- [3] Härmä A., Karjalainen M., Savioja L., Välimäki V., Laine U. K., and Huopaniemi J., "Frequency-warped signal processing for audio applications," *J. Audio Eng. Soc.*, vol. 48, no. 11, pp. 1011-1031, Nov. 2000.
- [4] Smith, J. O., and Abel, J. S. "The Bark Bilinear Transform," *Proc. IEEE ASSP Workshop*, New Paltz, 1995.
- [5] Karjalainen M., Härmä A., and Laine U.K., "Realizable Warped IIR Filters and Their Properties", *Proc. IEEE ICASSP-96*, Munich, 1996.
- [6] Härmä A., "Implementation of frequency-warped recursive filters," *EURASIP Signal Processing*, vol. 80, no. 3, pp. 543-548, March 2000.
- [7] Tolonen T., and Karjalainen M., "A computationally efficient multipitch analysis model," *IEEE Trans. Speech and Audio Proc.*, vol. 8, no. 6, pp. 708-716, 2000.
- [8] Karjalainen M., and Tolonen T., "Separation of speech signals using iterative multi-pitch analysis and prediction," in *Proc. EUROSPEECH'99*, vol. 5, pp. 2187-2190, Budapest, Sept. 5-9, 1999.
- [9] Karjalainen M., Altosaar T., and Vaino M., "Speech Synthesis Using Warped Linear Prediction and Neural Networks", *Proc. IEEE ICASSP-98*, Seattle, 1998.
- [10] Laine U. K., Karjalainen M., and Altosaar T., "Warped Linear Prediction (WLP) in Speech and Audio Processing," *Proc. IEEE ICASSP -94*, Adelaide, Australia, 1994.