Katholieke Universiteit Leuven

Departement Elektrotechniek

ESAT-SISTA/TR 07-28

Linear prediction of audio signals[1]

Toon van Waterschoot[2] [3] and Marc Moonen[2]

March 2007
Published in *Proceedings of Interspeech 2007,* Antwerp, Belgium, Aug. 27-31, 2007, pp. 518-521

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.be* in the directory *pub/sista/vanwaterschoot/reports/07-28.pdf*

[2]K.U.Leuven, Dept. of Electrical Engineering (ESAT), Research group SCD(SISTA), Kasteelpark Arenberg 10, 3001 Leuven, Belgium, Tel. +32 16 321927, Fax +32 16 321970, WWW: *http://homes.esat.kuleuven.be/∼tvanwate*. E-mail: *toon.vanwaterschoot@esat.kuleuven.be.*

[3]Toon van Waterschoot is a Research Assistant with the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of K.U.Leuven Research Council: CoE EF/05/006 Optimization in Engineering, the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office IUAP P6/04 (“Dynamical systems, control and optimization”, 2007-2011), the Concerted Research Action GOA-AMBioRICS, and IWT project 040803: “SMS4PA-II: Sound Management System for Public Address Systems”. The scientific responsibility is assumed by its authors.

Linear prediction of audio signals

Toon van Waterschoot and Marc Moonen

Dept. of Electrical Engineering (ESAT-SCD), Katholieke Universiteit Leuven, Leuven, Belgium

toon.vanwaterschoot@esat.kuleuven.be, marc.moonen@esat.kuleuven.be

Abstract

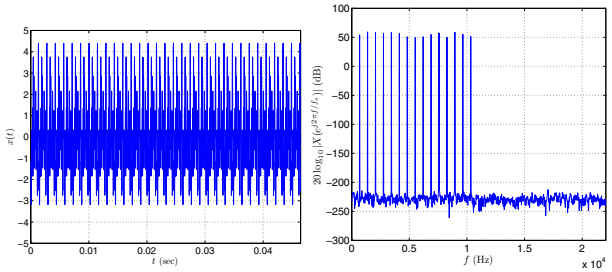
Linear prediction (LP) is a valuable tool for speech analysis and coding, due to the efficiency of the autoregressive model for speech signals. In audio analysis and coding, the sinusoidal model is much more popular, which is partly due to the poor performance of audio LP. By examining audio LP from a spectral estimation point of view, we observe that the distribution of the audio signal's dominant frequencies in the Nyquist interval is a critical factor determining LP performance. In this framework, we describe five existing alternative LP methods and illustrate how they all attempt to solve the observed frequency distribution problem.

Index Terms: linear prediction, spectral estimation, audio analysis, audio coding

1. Introduction

Linear prediction (LP) is a widely used and well understood technique for analysis and coding of speech signals [1]. Its success can be attributed to the insight in and efficient modelling of the speech generation process. The vocal tract can be modelled as a slowly time-varying, low-order all-pole filter, while the glottal excitation can be regarded either as a white noise sequence (for unvoiced sounds), or as an impulse train generated by periodic vibrations of the vocal chords (for voiced sounds). Using this so-called source-filter model, any speech fragment can be whitened using a cascade of a formant predictor for removing short-term correlation, and a pitch predictor for removing long-term correlation [2].

The source-filter model is much less popular in audio analysis than it is in speech analysis. First of all, the generation of musical sounds is highly dependent on the instruments used, hence it is hard to propose a generic audio generation model. Second, from a physical point of view, polyphonic audio signals should be analyzed using multiple source-filter models, which seems rather impractical. Finally, the transform-based audio coding approach has become immensely popular due to its succesful combination with perceptual models of human hearing [3]. Nevertheless, some audio coding algorithms still rely on LP (see [4] and references therein), which is then usually performed on a warped frequency scale. Also, in audio analysis applications other than coding, prediction error filters obtained with LP are used for whitening of audio signals, e.g., to produce robust and fast converging acoustic echo and feedback cancellers [5], [6].



(a) (b)

Figure 1: (a) Sum of sinusoids. (b) Sum of sinusoids spectrum.

Since many audio signals exhibit a large degree of tonality, i.e., their frequency spectrum is characterized by a finite number of dominating tonal components, it is useful to approach LP of audio signals in the frequency domain, i.e., from a spectral estimation point of view. Intuitively, it would be expected that performing LP using twice as many model parameters as the number of tonal components, leads to a signal estimate in which each of the spectral peaks is modeled with a complex conjugate pole pair close to (but inside) the unit circle. In practice, however, this does not seem to be the case, and very often a poor LP signal estimate is obtained. A first explanation for this is based on the observation that any realistic audio signal does not exactly match a sum of sinusoids, such that additive noise should be taken into account in the signal model. Whereas a sum of N sinusoids can be perfectly modelled using an AR($2N$) model, i.e., an autoregressive model of order $2N$, a sum of N sinusoids plus (white) noise should instead be modelled using an ARMA($2N$,$2N$) model, i.e., an autoregressive moving-average model with $2N$ MA and $2N$ AR coefficients [7],[8]. A second explanation is that LP is always performed using finite-sample data records, which turns a sum of sinusoids into a sum of narrowband signals. The smoothing effect so obtained is comparable to adding white noise to the sinusoids (i.e., the poles of the AR signal model are displaced towards the origin of the z-plane [8]), and again, an ARMA($2N$,$2N$) model should be used to obtain a correct signal estimate. In this paper, we provide a third explanation why an AR($2N$) model is not always appropriate for LP of a sum of N sinusoids, and as a consequence, why LP of audio signals often performs in a poor way. We will show that the distribution of the signal's tonal components in the Nyquist interval is an important factor w.r.t. LP performance. We will moreover illustrate how the existing solutions are in fact all different approaches to solve this very problem. More specifically, we will consider improved LP of audio signals by selective linear prediction [1], high-order modelling [8], pole-zero modelling [7],[9],[10], pitch prediction [2],[11], and

Toon van Waterschoot is a Research Assistant with the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of K.U.Leuven Research Council: CoE EF/05/006 Optimization in Engineering, the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office IUAP P6/04 ("Dynamical systems, control and optimization", 2007-2011), the Concerted Research Action GOA-AMBioRICS, and IWT project 040803: "SMS4PA-II: Sound Management System for Public Address Systems". The scientific responsibility is assumed by its authors.

Table 1: Spectral flatness measure (SFM) of LP residuals

LP method	SFM_R (dB)
LP (autocorrelation)	-6.92
LP (covariance)	-52.97
Pole-zero model	-1.19
High-order LP	-0.55
Pitch prediction	-15.37
Warped LP	-5.17
Selective LP	-1.76

frequency warping [4].

This paper is structured as follows. In Section 2, we illustrate how the distribution of the tonal components in the analyzed signal influences the performance of conventional LP. In Section 3, the aforementioned existing methods for improved LP are interpreted as being solutions to the problem highlighted in Section 2. Simulations results are presented throughout the text. Finally, Section 4 concludes the paper.

2. Conventional linear prediction: influence of frequency distribution

2.1. Preliminaries

We will only consider tonal audio signals, i.e., signals having a continuous spectrum containing a finite number of dominant frequencies. In this way, the majority of audio signals is covered, except for the class of percussive sounds. The performance of the LP methods described below, was evaluated for three audio fragments: a synthetic sum of sinusoids, a monophonic violin fragment, and a polyphonic piano fragment. Due to space limitations, only the results for the synthetic sum of sinusoids are described in this paper. We refer to [12] for the results using true monophonic and polyphonic audio signals.

A synthetic sum of $N = 15$ harmonic sinusoids with random, uniformly distributed amplitudes (in $[0, 1]$) and phases (in $[0, 2\pi]$) was generated as

$$x(t) = \sum_{n=1}^{N} A(n) \sin[n\omega_0 t + \phi(n)], \quad t = 1, ..., M. \qquad (1)$$

The normalized fundamental frequency $\omega_0 = 2\pi/64$ was chosen such that, at a sampling frequency $f_s = 44.1$ kHz, the fundamental frequency $f_0 \approx 689.1$ Hz is in the mid-range of musical notes (i.e., slightly lower than F5), and also such that the fundamental frequency and its harmonics are in the discrete set of frequencies at which the length-M discrete Fourier transform (DFT) is evaluated. Moreover, the pitch period $T_0 = 2\pi T_s/\omega_0 = 64T_s$ equals an integer number of sampling periods ($T_s = 1/f_s$), which allows us to clearly illustrate the effect of pitch prediction in Section 3.3, and T_0 is also an integer multiple of $2(N + 1)T_s$ to easily yield a selective LP method in Section 3.5. The fragment length was set to $M = 2048$ samples, corresponding to 46.4 ms at $f_s = 44.1$ kHz, i.e., 1/32 note at 161.5 beats per minute, such that it is musically relevant to assume stationarity over the entire fragment. The signal fragment and its spectrum are shown in Figs. 1(a) and 1(b), respectively.

The performance of the different LP methods will be evaluated both visually (by inspecting the frequency magnitude response of the prediction error filter (PEF), which should be the inverse of the sum of sinusoids spectrum), and using the spectral

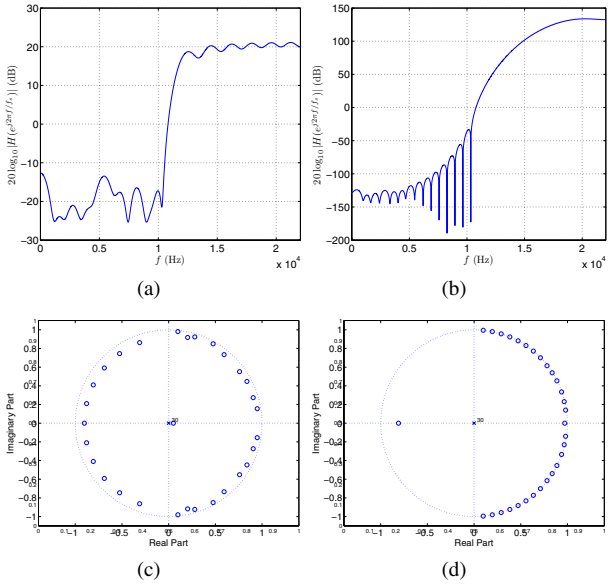


Figure 2: Conventional LP: (a) PEF frequency response (autocorrelation method), (b) PEF frequency response (covariance method). (c) PEF pole-zero plot (autocorrelation method), (d) PEF pole-zero plot (covariance method).

flatness measure (SFM) of the LP residual, defined as [8]

$$\text{SFM}_R(\text{dB}) = 10 \log_{10} \frac{\exp\left\{\frac{1}{M}\sum_{f=0}^{\tilde{f}_s} \ln\left[|R(f)|^2\right]\right\}}{\frac{1}{M}\sum_{f=0}^{\tilde{f}_s} |R(f)|^2}, \qquad (2)$$

where $R(f)$ represents the M-point FFT of the LP residual and $\tilde{f}_s = ((M-1)/M)f_s$ is the maximal FFT frequency. The SFM in dB is a negative real number, with SFM = 0 dB corresponding to a flat spectrum. The residual SFM of the LP methods described below is given in Table 1. Note that the original sum of sinusoids has an SFM equal to -39.93 dB.

2.2. Conventional LP

If we perform a $2N$-th order conventional LP on the above sum of sinusoids fragment, using the autocorrelation method and the Levinson-Durbin recursion for estimating the AR coefficients, we obtain a PEF with a frequency response and a pole-zero plot as shown in Figs. 2(a) and 2(c). Instead of cancelling each of the N sinusoids, the PEF has $N/2$ complex conjugate zero pairs in the sum of sinusoids' frequency region, and another $N/2$ complex conjugate zero pairs in the higher half of the Nyquist interval. If we use the covariance method instead of the autocorrelation method, we nearly achieve exact cancellation of the sinusoids, as shown in Figs. 2(b) and 2(d). However, since the fundamental frequency and its harmonics are all located in the lower half of the Nyquist interval, such a pole-zero configuration produces an undesirably large amplification in the high-frequency region. This observation can be extrapolated to the majority of audio signals since their fundamental frequency is typically below 1 kHz, hence when sampled at $f_s = 44.1$ kHz, the signal's most important harmonics lie in the lower half of the Nyquist interval. From the SFM values in Table 1, it can be understood that the covariance LP method does not lead to a spectrally flat residual, whereas the autocorrelation LP method sacrifices part of its zeros to achieve spectral flatness in the high-frequency region. We should also note that when noise is added

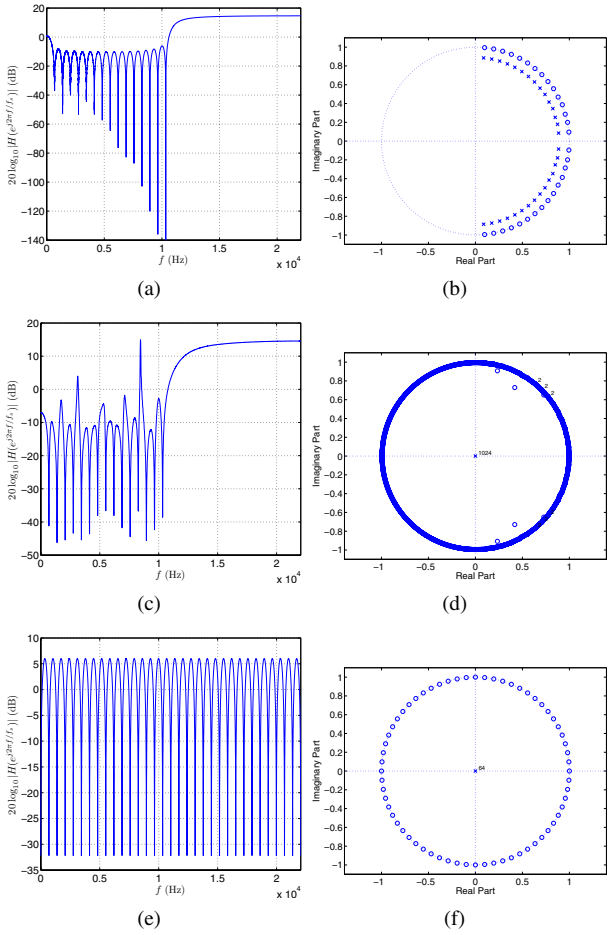


Figure 3: (a) Pole-zero model: PEF frequency response. (b) Pole-zero model: PEF pole-zero plot. (c) High-order LP: frequency response. (d) High-order LP: pole-zero plot. (e) Pitch prediction: PEF frequency response. (f) Pitch prediction: PEF pole-zero plot.

to the synthetic sum of sinusoids signal, both methods perform in a similar way, and not much different from the autocorrelation method in the noiseless case.

3. Methods for improved linear prediction

In this section, we present five existing alternative methods for LP of audio signals, and we illustrate how all these methods attempt to compensate for the effect described in Section 2.2. We will mainly focus on the alternative LP models, and not on how the model parameters can be estimated.

3.1. Pole-zero modelling

True audio signals can generally not be exactly modelled as a sum of sinusoids, hence a noise term should be added to account for imperfections of the model. It is well-known that a sum of N sinusoids plus white noise can be exactly modelled using an ARMA$(2N, 2N)$ model with equal coefficients in the AR and MA parts [7]-[9], however, the coefficients of the PEF denominator are usually scaled using a so-called pole radius [9],[10]. Figs. 3(a) and 3(b) show the frequency response and pole-zero plot of a pole-zero PEF with zeros on the unit circle at angles correponding to the sinusoid frequencies $\pm n\omega_0, n = 1, \ldots, N$ and poles inside the unit circle at radius $\rho = 0.9$ and at the same angles as the zeros. This pole-zero model produces N notches at the desired frequencies, while keeping the remainder of the residual spectrum relatively flat. Moreover, by constraining the zeros to lie on the unit circle, the coefficients of the MA part are mirror symmetric, such that the model is made up of only $N+1$ distinct parameters [10].

3.2. High-order AR modelling

High-order AR modelling was proposed earlier as an alternative to ARMA modelling of sinusoidal signals buried in noise [8]. Performing an 1024-th order LP to the sum of sinusoids fragment, yields the PEF with frequency response and pole-zero plot shown in Figs. 3(c) and 3(d). It can be seen that the PEF frequency response is very close to the inverse of the sum of sinusoids spectrum. Examining the distribution of the zeros in the complex plane reveals that this method produces approximately $1024 - 2N$ zeros, lying on and nearly equally spaced around the unit circle (to provide overall spectral flatness of the frequency response), and $2N$ additional zeros at the sinusoid frequencies $\pm n\omega_0, n = 1, \ldots, N$ (to provide the notch characteristics). The main disadvantage of this method is its high model complexity.

3.3. Pitch prediction

Periodic signals with a pitch period T_0 corresponding to an integer number of sampling periods T_s can be perfectly predicted using a one-tap pitch predictor,

$$P(z) = 1 - z^{-T_0/T_s}. \tag{3}$$

As shown in Figs. 3(e) and 3(f), this corresponds to an all-zero comb filter, featuring T_0/T_s zeros lying on and equally spaced around the unit circle. Hence, the pitch predictor produces an extremely non-smooth residual spectrum by also attempting to cancel harmonic frequencies which are not present in the input signal. Obviously, the pitch period will generally not be an integer multiple of the sampling period, such that the use of a multi-tap pitch predictor is required for interpolation (see, e.g., [2]), or a fractional delay filter should be used [11].

3.4. Warped linear prediction

Warped linear prediction (WLP) is probably the most well-known technique for LP of audio signals [4]. In WLP, the input signal undergoes a frequency transformation before LP is performed, with the aim of enhancing the frequency resolution in certain frequency regions. The frequency transformation is usually chosen such that it approximates the Bark auditory scale [13]. Fig. 4(a) shows the spectrum of the frequency-warped sum of sinusoids, and Figs. 4(c) and 4(e) illustrate the PEF frequency response and pole-zero plot after a $2N$-th order WLP is performed. The sinusoid frequencies are now spread around the unit circle, thereby enhancing the LP performance. It can be seen that for monophonic, harmonic signals, this type of frequency mapping is not ideal, since the frequency spreading is not uniform, hence lower resolution is obtained in the high-frequency region. The WLP method is however better suited for polyphonic signals [12].

3.5. Selective linear prediction

Compared to WLP, a linear frequency mapping is more useful for monophonic, harmonic signals, since it preserves the harmonic relation between the sinusoid frequencies. The ideal

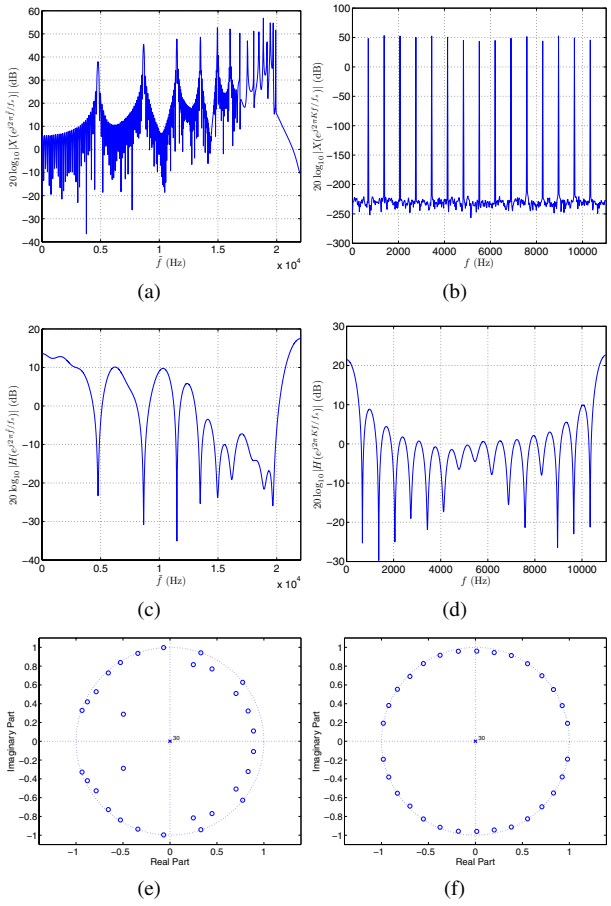


Figure 4: (a) Warped sum of sinusoids spectrum. (b) Downsampled sum of sinusoids spectrum. (c) Warped LP: PEF frequency response. (d) Selective LP: PEF frequency response. (e) Warped LP: PEF pole-zero plot. (f) Selective LP: PEF pole-zero plot.

mapping for a sum of N sinusoids in fact corresponds to a downsampling operation with downsampling factor $K = \pi/(N+1)\omega_0$ [12]. This corresponds to a special case of so-called selective linear prediction (SLP) [1], and causes the LP to take only the frequency interval $\omega \in [0, (N+1)\omega_0]$ into account. However, an integer downsampling factor is only obtained when the pitch period T_0 is an integer multiple of $2(N+1)T_s$, which is the case in our example, but not in general. The spectrum of the sum of sinusoids, downsampled with a factor $K = 3$, is shown in Fig. 4(b), and the PEF resulting from a $2N$-th order SLP has a frequency response and pole-zero plot as shown in Figs. 4(d) and 4(f). The SLP performs relatively well, but obviously the ideal frequency mapping is highly signal-dependent, and non-integer downsampling is required in general.

4. Conclusion

We have illustrated that for LP of tonal signals, the distribution of the signal's tonal components in the Nyquist interval has a large impact on LP performance. Five alternative LP methods were reviewed and evaluated by examining the PEF frequency characteristics and the spectral flatness of the corresponding LP residual. The pole-zero model, the high-order LP, and the selective LP were shown to produce a spectrally much flatter residual than the conventional LP. However, the high-order LP is computationally much more demanding, and the selective LP relies on a frequency mapping which is highly signal-dependent. Hence the pole-zero model, corresponding to an ARMA$(2N, 2N)$ signal model, seems to be the best alternative for LP of a signal containing N tonal components.

5. References

[1] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–578, Apr. 1975.

[2] R. P. Ramachandran and P. Kabal, "Pitch prediction filters in speech coding," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-37, no. 4, pp. 467–478, Apr. 1989.

[3] K. Brandenburg and G. Stoll, "ISO/MPEG-1 audio: A generic standard for coding of high-quality digital audio," *Journal of the Audio Eng. Soc.*, vol. 42, no. 10, pp. 780–792, Oct. 1994.

[4] A. Härmä and U. K. Laine, "A comparison of warped and conventional linear predictive coding," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 579–588, July 2001.

[5] T. van Waterschoot, G. Rombouts, P. Verhoeve, and M. Moonen, "Double-talk-robust prediction error identification algorithms for acoustic echo cancellation," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 846–858, Mar. 2007.

[6] G. Rombouts, T. van Waterschoot, K. Struyve, and M. Moonen, "Acoustic feedback suppression for long acoustic paths using a nonstationary source model," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3426–3434, Sept. 2006.

[7] W. J. Fitzgerald and R. Geere, "Class of constrained ARMA models for line enhancement using real-time QR implementation," *IEE Electronics Lett.*, vol. 27, no. 24, pp. 2230–2231, Nov. 1991.

[8] S. M. Kay, "The effects of noise on the autoregressive spectral estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, no. 5, pp. 478–485, Oct. 1979.

[9] D. V. B. Rao and S.-Y. Kung, "Adaptive notch filter for the retrieval of sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 4, pp. 791–802, Aug. 1984.

[10] A. Nehorai, "A minimal parameter adaptive notch filter with constrained poles and zeros," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 4, pp. 983–996, Aug. 1985.

[11] P. Kroon and B. S. Atal, "Pitch predictors with high temporal resolution," in *Proc. 1990 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '90)*, vol. 2, Albuquerque, NM, USA, Apr. 1990, pp. 661–664.

[12] T. van Waterschoot and M. Moonen, "Comparison of linear prediction models for audio signals," *IEEE Trans. Audio Speech Lang. Process.*, submitted for publication. [Online]. Available: ftp://ftp.esat.kuleuven.be/pub/sista/vanwaterschoot/abstracts/07-29.htm%l

[13] J. O. Smith and J. S. Abel, "Bark and ERB bilinear transforms," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 6, pp. 697–708, Nov. 1999.