DETECTION OF CLICKS IN AUDIO SIGNALS USING WARPED LINEAR PREDICTION

Paulo Esquef[1]*, *Matti Karjalainen*[1]

[1]Helsinki University of Technology
Lab. of Acoustics and Audio Signal Processing, P.O.Box 3000, FIN-02015 HUT, Espoo, Finland
esquef@acoustics.hut.fi

Vesa Välimäki[2,1]

[2]Tampere University of Technology
Pori School of Technology and Economics, P.O.Box 300, FIN-28101 Pori, Finland
vesa.valimaki@pori.tut.fi

ABSTRACT

In this work *warped linear prediction* (WLP) is applied to a model-based method to detect impulsive disturbances in audio signals. According to simulations performed on artificially corrupted audio signals the adoption of negative values for the warping factor favors the click detection scheme. As a consequence, for equal levels of missing (false) detection the WLP-based scheme yields consistently lower percentage of false (missing) detection than the conventional method.

1. INTRODUCTION

Frequency warping techniques have been applied successfully to several audio applications [1, 2]. In this work, WLP is applied to a model-based method of audio declicking (impulsive noise removal) and its performance is confronted against the conventional method.

In one of the most popular model-based methods for audio de-clicking, short-time signal frames of the uncorrupted audio signal are modeled as autoregressive (AR) processes, and linear prediction is used to detect the corrupting clicks in the signal [3, 4]. The modeling performance plays an important role on the detection scheme, and thus it is worth investigating possible benefits of using WLP for this task.

According to simulations carried out in this work, the use of WLP in the model-based click detection is advantageous. Its main benefit is the possibility to improve the prediction gain, which has a close connection with the click detection performance. Simulations on real musical signals artificially corrupted show that the percentage of missing detection can be reduced when a suitable value for the warping factor is adopted.

This paper is organized as follows. In Section 2, the basic principles of the AR-based technique for impulsive noise detection are reviewed. In Section 3, the basic concepts of frequency warping and WLP are addressed, and the performance of the WLP-based click detection scheme is evaluated for several values of the warping factor. Experiments and results are described in Section 4. Conclusions are drawn in Section 5.

*The work of Paulo Esquef has been financed by the Brazilian National Council for Scientific and Technological Development (CNPq-Brazil) and by the Academy of Finland project "Technology for Audio and Speech Processing".

2. MODEL-BASED AUDIO DE-CLICKING

Impulsive disturbances or clicks can be described as localized discontinuities of short duration (typically less than 1 ms) that randomly corrupt an underlying signal [4].

Digital signal processing techniques for de-clicking purposes can be in general separated in two stages: detection of clicks and signal reconstruction. Usually, both the detection and the reconstruction stages employ model-based approaches within block-processing schemes.

2.1. Audio Modeling

Consider a sequence containing N samples of the corrupted signal, modeled as $y(k) = x(k)+d(k)$, where $d(k)$ is the noise sequence, and $x(k)$ is the uncorrupted signal, which is modeled as a p^{th}-order AR process defined by

$$x(k) = \sum_{n=1}^{p} a(n)x(k-n) + e(k), \quad k = p, \cdots, N-1, \tag{1}$$

where $a(n)$ are the model parameters and $e(k)$ is the excitation signal or the prediction error sequence.

The model parameters can be estimated by minimizing the prediction error energy, for instance, through either the covariance or autocorrelation methods. As the parameter estimation is performed over the corrupted signal, a biased estimate is inevitable.

2.2. Detection Stage

The basic steps of the detection stage consist of estimating the model parameters, inverse filtering of the noisy signal, and applying a selection criterion over the excitation sequence to detect the corrupting clicks.

The excitation signal obtained from a corrupted signal block can be written as

$$e(k) = e_x(k) + d(k) - \sum_{j=1}^{p} d(k-j)a(j), \tag{2}$$

0-7803-7503-3/02/$17.00 ©2002 IEEE

where $e_x(k)$ is the excitation term associated with the uncorrupted signal, and $d(k)$ is the impulsive noise sequence itself. However, the last term indicates that the clicks are also spread in the excitation signal. For the click detection purpose, the lower the variance of $e_x(k)$ and the less prominent the spreading term, the better the click detection performance.

One possible criterion to detect the clicks as well as to determine their location in time [3, 4] consists of considering corrupted all those samples in the signal that correspond to excitation samples whose magnitude exceeds a given threshold ϵ. The value of the threshold can be obtained as $\epsilon = K\,\hat{\sigma}_{e_x}$, where $\hat{\sigma}_{e_x}$ is an estimate of the excitation standard-deviation and K is a gain factor. The value of $\hat{\sigma}_{e_x}$ can be estimated from the noisy excitation sequence by applying a median estimator [3]. The value of K is settled empirically to control the tradeoff between percentage of false and missing detection. Its value is not changed along the block-processing scheme.

The performance of threshold-based methods for click detection purposes can be improved by other means, for instance, the inclusion of a matched filtering procedure after the inverse filtering, the use of two-sided and extended linear prediction formulations [3], and the adoption of a double-threshold detection scheme [5].

2.3. Reconstruction Stage

The last step of the de-clicking process consists of replacing the corrupted samples by others that resemble the audio behavior in the neighborhood of the disturbances.

In this work, the Least Squares AR-based interpolator [4] was used. It estimates the new samples by minimizing the prediction error energy with respect to the corrupted samples. The method is well suited for interpolation of moderate gaps (up to 100 samples at 44100 Hz) in the audio signal. In this sense, some level of false alarm in the click detection is not very harmful and can be tolerated if it yields short gaps, since one can count on the reconstruction stage to recover the signal. It is more important to aim at a low missing detection rate.

3. WLP-BASED CLICK DETECTION

3.1. Warped Linear Prediction

Parametric representations of signals such as the linear prediction coefficients can be carried out in a warped frequency domain. The WLP uses the bilinear conformal mapping [6]. In this case, the unit delays of the analysis FIR filter are replaced by first-order allpass filters. The modified delay element is defined as

$$\tilde{z}^{-1} = D(z) = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}}, \quad |\lambda| < 1, \tag{3}$$

where λ is the warping factor. The group delay of $D(z)$ varies as a function of frequency. Therefore, it is possible to attain a non-uniform frequency resolution by properly setting the value of λ. Within this structure, the choice of positive values of λ leads to an increased resolution at low frequencies whereas negative values of λ yield an increased resolution at high frequencies.

Table 1. Prediction gain G_p for different values of λ.

λ	-0.8	-0.4	0.0	0.4	0.8
G_p	35.51	27.49	22.37	17.30	11.04

The description of the algorithms to compute the WLP coefficients as well as the implementation of the analysis (WFIR) and synthesis (WIIR) filters can be found in [7]. Furthermore, the use of frequency warping filtering techniques are computationally more demanding than the standard ones. However, an extra cost is not critical for audio restoration applications, since real-time processing is usually not required.

3.2. WLP-based Click Detection

In the WLP-based detection scheme the conventional estimation of the AR model parameters is replaced by a WLP estimator, and the inverse filtering procedure is carried out using a WFIR structure to obtain an excitation with temporal resolution equal to that of the analyzed signal. The effects of this scheme on the prediction gain and the spreading of clicks are investigated next.

3.2.1. Prediction Gain

There are many ways to assess the modeling performance of a linear prediction scheme. For click detection purposes the goal of the inverse filtering is to produce a high contrast between the prediction error of the corrupted samples and that of the uncorrupted samples. Therefore, instead of employing a spectral flatness measure over the excitation, it is better to evaluate the modeling performance via the relative improvement in the noise-to-signal ratio due to the inverse filtering. If the spreading term in (2) is not considered, this can be measured via the prediction gain, defined by

$$G_p = 10 \log_{10} \left(\frac{\sum_{n=1}^{N} |x(n)|^2}{\sum_{n=1}^{N} |e_x(n)|^2} \right), \tag{4}$$

where $x(i)$ is the uncorrupted signal and $e_x(i)$ its corresponding excitation within a given signal block.

In order to evaluate how the WLP can favor the detection scheme under a real situation, a mean G_p was computed as a function of the warping factor λ for a 5 s segment of orchestral music artificially corrupted by clicks. This signal was sampled at 44.1 kHz and segmented in 215 frames of 1024 samples. For each frame, 40^{th}-order WLP prediction filters were estimated for values of λ sampled from -0.8 to 0.8 in steps of 0.4. As the location of the clicks was known beforehand, the corrupted samples were discarded from the G_p computation.

The obtained result of the mean G_p over 215 frames is shown in Table 1, from where it can be verified that, on the average, the prediction gain tends to decrease as λ is increased. Note that the model parameters used to compute $e_x(i)$ in Eq. 4 were estimated from noisy data. Therefore, although not explicitly shown in Eq. 4, the values of G_p are influenced by the presence of clicks in the signal.

According to Table 1, the more close to -1 the value of λ is set the higher the prediction becomes. However, a high value of G_p does not necessarily mean a better signal modeling performance. It can simply reflect how well the signal energy is reduced after the inverse filtering [7] (p. 94).

The previously described behavior lays on the fact that the excitation produced by the WFIR inverse filter does not have a flat power spectrum. Actually, the excitation spectrum has a tilt given by the squared magnitude of $W(z) = \sqrt{(1-\lambda^2)}/(1-\lambda z^{-1})$ [6, 7]. Therefore, when negative values of λ are adopted the spectrum assumes a highpass filter characteristic, whereas a lowpass filter profile is obtained for positive values of λ.

The spectral tilt observed in the excitation can be corrected by prefiltering the signal through $W^{-1}(z)$. However, such a correction serves nothing to threshold-based click detection. On the contrary, it is desirable to detect clicks through an excitation which has its high-frequency components emphasized. This holds true when employing negative values of λ.

3.2.2. *Spreading of Clicks*

As seen in Section 2.2, the spreading of clicks is determined by the impulse response of the inverse filter. In the WLP-based scheme, the inverse filter is WFIR with infinite impulse response, since its internal elements, $D(z)$, are first-order recursive allpass filters. It can be shown [8] that the energy carried by the impulse response of $D(z)$ is more concentrated in its initial samples when $|\lambda|$ approaches 0. Thus, it is plausible to expect longer impulse responses for the WFIR filter, and consequently more pronounced spreading effects, when $|\lambda|$ is set close to 1.

The spread of the clicks affects mainly the determination of its length when running a threshold-based detection scheme. However, the beginning of the clicks is still well defined, since WLP implies a one-sided forward prediction. Therefore, an effective way to overcome the pronounced spreading of the clicks is to 1) compute an additional excitation by inverse filtering a time-reversed version of signal frame; 2) obtain two sets containing the indices associated with the corrupted samples, one for each excitation; 3) and then, take only those indices that are common to both sets. This strategy corresponds to estimating the beginning of the clicks through the forward inverse-filtered signal and the end point using the backward inverse-filtered signal.

4. EXPERIMENTS AND RESULTS

The signals addressed in this work are real musical extracts which were artificially corrupted by impulsive noise to allow the use of objective measures, e.g., the missing detection percentage (MDP) and false detection percentage (FDP).

Usually, both the MDP and FDP vary strongly according to the values of the parameters employed in the click detection algorithm. Additionally, the quality of the signal reconstruction is not considered in the previous measures. The WLP is only applied to the detection stage as the main

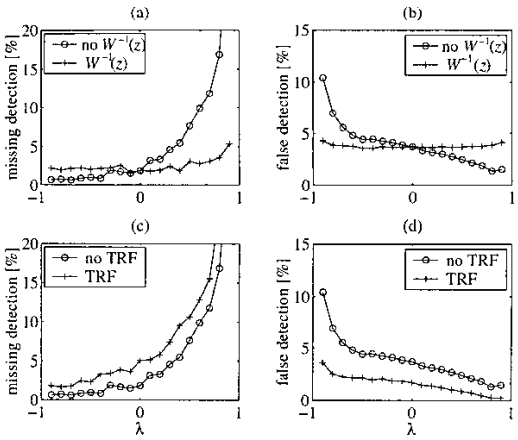

Fig. 1. MDP and FDP in the WLP-based click detection as a function of the warping factor measured for S1. Plots (a) and (b) illustrate the effect of the prefiltering whereas plots (c) and (d) illustrate the effect of the TRF.

intention here is to evaluate its effects on the detection performance.

Two test signals S1 and S2 (monaural, 44.1 kHz, 16 bits, duration about 10 s) containing excerpts of orchestral music are used in the simulations. The impulsive noise sequences are taken as the difference between other originally corrupted signals and their restored versions. The percentage of corrupted samples in S1 and S2 are approximately 0.6% and 5%, respectively.

The evaluation strategy consists of first setting the parameters of the detection method in order to produce a satisfactory restored result. In this initial calibration procedure the value of λ was set to zero, which is equivalent to employing conventional linear prediction. The model order, the threshold gain, and the frame length (see Section 2) were set to $p = 40$, $K = 5$, and $N = 1024$, respectively. At this stage, the time-reversed filtering (TRF) (see Section 3.2.2) was not used. The interpolation algorithm employed in the reconstruction stage was the Least Squares AR-based, as described in [4].

Now, to assess the effect of the WLP on the click detection performance the values of K and p are frozen, and the value of λ is varied from -0.9 to 0.9 in steps of 0.1. For the sake of clarity, the consequences of including or not the prefiltering, $W^{-1}(z)$, and the TRF resources over the MDP and FDP rates are depicted separately in Fig. 1. It can be verified that, for negative values of λ, the lack of prefiltering yields a reduction in MDP whereas FDP is increased. On the contrary, the employment of the TRF produces a small increase in MDP followed by a reduction in FDP.

Additional simulations also showed that the MDP and FDP measures are not sensitive to the choice of the model order, at least within the range between 10 and 80. Similar behavior was observed for S2, although the overall performance is worse than that of S1, due to its much higher percentage of corrupted samples.

According to the plots shown in Fig. 1, it is impossible to choose an optimal solution for λ which minimizes

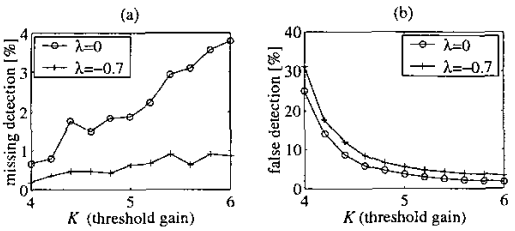

Fig. 2. Percentage of (a) missing and (b) false detection as functions of the threshold gain.

both the MDP and the FDP. However, considering that it is preferable to compromise the minimization of FDP in favor of MDP, the best option is to set a negative λ for the WLP and exclude the prefiltering stage.

A good strategy to confront the performance of the click detection using conventional linear prediction, $\lambda = 0$, against WLP with $\lambda = -0.7$, but without prefiltering, is to fix all other parameters and vary the threshold gain, K, which has direct impact on both missing and false detection percentage. The results obtained when evaluating the signal S1 are shown in Fig. 2. As an example, to achieve MDP below 1% when employing $\lambda = 0$, it is necessary to set $K \leq 4.2$, which implies a FDP of about 14%. On the other hand, the same requirement is satisfied when using $\lambda = -0.7$ by setting $K = 6$, and in this case, the attained FDP is 3.4%. If the detection stage includes the TRF scheme, the use of $\lambda = -0.7$ is still advantageous since it yields lower levels of false alarm in spite of higher MDP. When adjusting K to achieve a MDP below 2% for both $\lambda = 0.0$ and $\lambda = -0.7$ the resulting FDP are 7.2% and 1.8%, respectively. It can be concluded then that the use of WLP with $\lambda = -0.7$ yields as low MDP as the case with $\lambda = 0.0$, but achieves a lower level of FDP.

It is worth noticing that the possibility to lower the MDP and the FDP surely reflects a better performance of the detection scheme. However, its relation with the perceptual quality of the restored signals can be highly nonlinear. For instance, reducing an already small MDP only produces subtle improvements which may be hard to perceive. According to informal listening tests, the perceptual improvement attained by using WLP with negative λ in the click detection stage can be perceived at some specific passages, e.g., during the *crescendo* in S2. Audio samples are available at URL:
http://www.acoustics.hut.fi/publications/papers/dsp2002-declick/

5. CONCLUSIONS

This paper proposed the use of warped linear prediction in a model-based click detection method. It was found that by properly setting the warping factor, λ, the spectrum of the excitation is emphasized at higher frequencies favoring the threshold-based click detection. The desirable increase in the prediction gain can be attained by reducing the value of λ at the cost of a more pronounced spreading of clicks. To overcome the latter, an additional inverse filtering using a time-reversed version of the signal frame was proposed. Simulations were performed on real musical signals, artificially corrupted by impulsive noise. The results show that the missing detection decreases when the value of λ is decreased, although the opposite behavior is observed for the percentage of false detection. A case study confronting the performance of detection scheme using WLP ($\lambda = -0.7$) against the conventional linear prediction ($\lambda = 0$) showed that, in the former case, it is possible to achieve equal levels of missing but at a lower percentage of false detection and vice-versa. Finally, it is interesting to notice that the use of WLP in audio applications is usually intended to perform signal analysis based on auditory modeling, which can be achieved via WLP by adopting a positive value for λ [9, 2]. However, this is not the case for the click detection scheme presented in this paper, which shows that signal modeling with $\lambda < 0$ is better suited for detection of short clicks.

6. REFERENCES

[1] M. Karjalainen, A. Härmä, U. K. Laine, and J. Huopaniemi, "Warped filters and their audio applications," in *Proc. IEEE WASPAA'97*, New Paltz, New York, 1997.

[2] A. Härmä, M. Karjalainen, V. Välimäki, L. Savioja, U. K. Laine, and J. Huopaniemi, "Frequency-warped signal processing for audio applications," *J. Audio Eng. Soc.*, vol. 48, no. 11, pp. 1011–1031, Nov. 2000.

[3] S. V. Vaseghi and P. J. W. Rayner, "Detection and suppression of impulsive noise in speech communication systems," *IEE Proceedings*, vol. 137, no. 1, pp. 38–46, Feb. 1990.

[4] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration — A Statistical Model Based Approach*, Springer-Verlag, London, UK, 1998.

[5] P. A. A. Esquef, L. W. P. Biscainho, P. S. R. Diniz, and F. P. Freeland, "A double-threshold-based approach to impulsive noise detection in audio signals," in *Proc. EUSIPCO'00*, Tampere, Finland, Sept. 2000, vol. 4, pp. 2041–2044.

[6] H. W. Strube, "Linear prediction on a warped frequency scale," *J. Acoust. Soc. Am.*, vol. 68, no. 4, pp. 1071–1076, Oct. 1980.

[7] A. Härmä, "Audio coding with warped predictive techniques," M.S. thesis, Helsinki Univ. of Technology, Espoo, Finland, Jan. 1998, Available at URL: http://www.acoustics.hut.fi/publications.

[8] T. I. Laakso and V. Välimäki, "Energy-based effective length of the impulse response of a recursive filter," *IEEE Trans. Instrum. Meas.*, vol. 48, no. 1, pp. 7–17, Feb. 1999.

[9] 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.