Transients Detection and Segmentation in Audio Signals

Francois Xavier Nsabimana, Udo Zölzer
Helmut-Schmidt-University / University of the Federal Armed Forces Hamburg, Germany
Department of Signal Processing and Communications
Email: fransa, udo.zoelzer@hsu-hh.de

Abstract

In this paper, we present a time domain frame based approach for transient detection and modelling of audio signals. Most audio signals are mainly composed of two parts: a deterministic and a stochastic part. The deterministic part of an audio signal consists of sinusoids, while noise and transients represent the stochastic part. Although the Spectral Modelling Synthesis approach proposed by Smith and Serra improves the originally Sinusoidal Modelling proposed by McAulay and Quatieri by taking into account the stochastic part, it does not make any differentiation between noise and transients. Since transients do not fit well into the proposed noise model, they therefore need their own model. Our time domain frame based approach is using linear prediction to calculate the prediction error of an audio signal. Since sudden changes in audio signals, like transients, remain unpredictable, the prediction error will show transients more accentuated in transient parts of the audio signal. Based on that assumption, we build an adaptive threshold based on envelope estimation of the obtained prediction error. The threshold function is kept constant in the transient area and follows the signal elsewhere. If a complete transient or a beginning of transient is detected in one frame, it will automatically be separated. A three components model is built and transients are thus left to their own model. With that technique we succeed in detecting and separating several transients in different types of audio signals.

Introduction

Recent audio signal processing techniques are making some differentiation between deterministic and stochastic parts on the one hand, and between noise and transients on the other hand [1]. A model which is suitable for sinusoids may not be really suitable for noise and transients. Although noise and transients both represent the stochastic part of an audio signal, they do not fit well in the same model. A three components model is therefore required. In [2] a model for sinusoids with time-varying amplitudes, phases and harmonic frequencies has been presented. In [3], the sinusoidal model (SM) proposed in [2] was extended with a noise model based on residual approximation. The Spectral Modelling Synthesis (SMS) presented in [3] gives good results when applied to audio signals only composed of sinusoids and noise. But once transients occur in an audio signal, they will then appear in the residual signal. This will thus raise the spectral envelope of the noise during a residual approximation, yielding a synthesized signal with artefacts. To avoid this, a pre-processing step is required to first separate the transient's contribution.

Several methods to detect and model transients have already been presented. In [1] T.S. Verma and T.H.Y. Meng have proposed an improvement of the SMS model. But the transient modelling synthesis approach (TMS) presented in [1] requires a DCT block length of about 1 sec in order to accurately detect transients. In that model, transients occurring near the end of a DCT block were not detected and thus not modeled. In [4] a slightly modified TMS model was presented, where it has been shown how the TMS could succeed in detecting all transients within a DCT block.

In this paper, we present a time domain frame based approach for transient detection in audio signals based on linear prediction within the frame. In Section 2, we will introduce our proposed model, Section 3 will show results of the model applied to audio signals. We will finally close with a conclusion and outlook in Section 4.

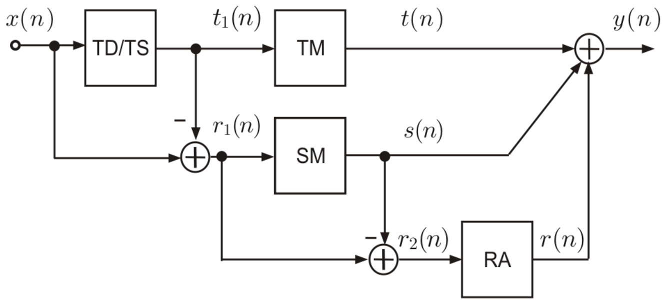

Figure 1: *Transients + Sinusoids + Noise Model. TD/TS (Transient Detection / Transient Separation), TM (Transient Modelling), SM (Sinusoidal Modelling), RA (Residual approximation). s(n): sinusoids, t(n): transients, r(n): noise.*

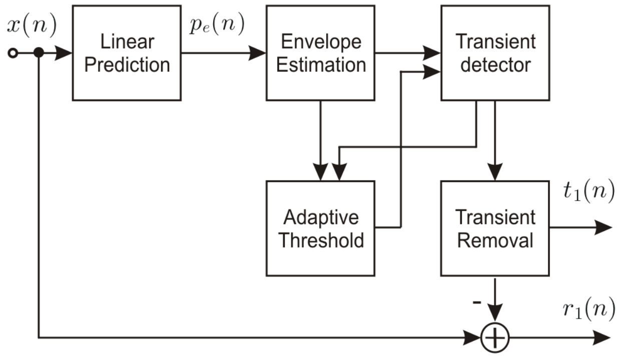

Figure 2: *TD/TS: Transient Detection & Separation approach. $t_1(n)$: detected transients, $r_1(n)$: first residual.*

The proposed Model

In Fig. 1 the whole system is presented, while in Fig. 2 the pre-processing steps for the transient model are detailed. Transients can be classified into two categories: visible and hidden transients. The main difficulty does not consist of detecting visible transients, but those of small energy. A good transients detector should thus reveal the presence of hidden transients and then emphasize them. In order to successfuly detect both kinds of transients, we need to apply a filter which should absorb most of the audio signals (sinusoids and noise) energy leaving transients unchanged. Since sudden changes in audio signals, like transients, remain unpredictable, the prediction error will then more accentuate transients in the audio signal.
The main steps for transient detection as presented in Fig. 2 can be described as follows. An input signal $x(n)$ is decomposed into short frames of 1024 samples with an overlap of 512 samples. Linear prediction is then applied in each frame to reveal the transients locations. A prediction filter with model order $p = 10$ is sufficient for our application. A suitable envelope estimator is applied to the prediction error to build the threshold. The threshold function will be kept constant in transient area and follows the signal elsewhere. This is done by comparing the actual value of the envelope with the weighted mean value of the envelope from the previous frame. If the value of the envelope is higher than the weighted mean value, the threshold function is kept constant equal to the non-weighted mean value of the envelope from the previous frame. A binary sequence is then set to one at that index. The index where the obtained binary sequence is equal to one will exactly correspond to index where transients are occurring in the current frame (see Fig. 3 lower right). Since transients which occur complete or truncated in the previous frame, could again appear as truncated or complete in the current frame, a strategy has been developed to avoid multiple transient detection in the overlapping zone.
If the current frame is free of transients, a frame with zeros will be generated and no transient modelling is performed for that frame. But if transients are identified, they will automatically be separated. The gaps left when transients are removed, will then be restored applying forward and backward extrapolation. A three components model is built and transients, sinusoids and noise are thus left to their respective model.

Application to Audio signals

Fig. 3 shows the original castanets signal and separated transients on the left and the prediction error and located transient areas on the right. Fig. 4 shows the results of the same technique applied to a polyphonic signal.

Conclusion

The transients detection approach as presented in this paper succeeds in detecting and separating several transients in audio signals. We have shown that performing linear prediction in each frame helps to reveal and emphasize visible and hidden transients. Applying an adaptive threshold is helpful to detect small energy transients. In future work, a three components model will be implemented for better signal reconstruction.

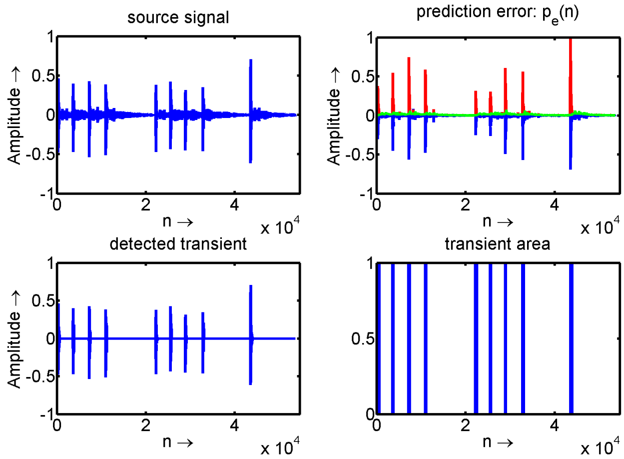

Figure 3: *Transients detection in Castanets sound file.*

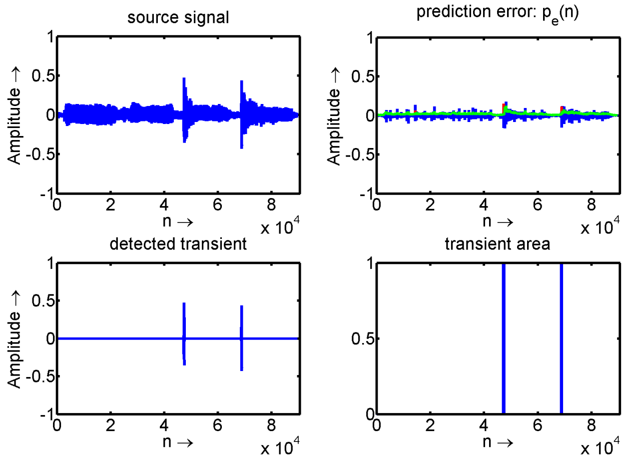

Figure 4: *Transients detection in ABBA sound file.*

References

[1] T. Verma and T.H.Y. Meng. Extending spectral modeling synthesis with transient modeling synthesis. In *Computer Music Journal, 24(2):47–59*, 2000.

[2] R. McAulay and T. F. Quatieri. Speech analysis/synthesis based on a sinusoidal speech model. *IEEE Transactions on Acoustics, Speech and Signal Processing*, pages 43:744–754, 1986.

[3] X. Serra and J.O. Smith. Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. In *Computer Music Journal, 14(4):14-24*, 1990.

[4] F.X. Nsabimana and U. Zölzer. Analysis/synthesis of transients in audio signals. DAGA´06, Braunschweig, 20-23 March 2006.