Analysis/Synthesis of Transients in Audio Signals

Francois Xavier Nsabimana[1], Udo Zölzer[2]
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

Sinusoidal modeling (SM) is a technique, which represents an audio signal as a sum of sinusoids with time varying frequency, amplitude and phase. An extension of this spectral modeling takes into account that the signal is composed of a deterministic (the sinusoidal part of the signal) and a stochastic part. In this model the non-periodic components (stochastic part) will be obtained by filtering white noise with the spectral envelope of a residual signal (error signal). The Sinusoids + Transients + Noise model (STN) to be presented in this paper splits the stochastic part of the signal into noise and transients. This model proposed by T.Verma and T.H.Y.Meng uses the duality between time and frequency domain to track transients. Since transients are impulsive signals in the time domain, they will be mapped to sinusoidal signals in the frequency domain. Such a mapping will be provided by a discrete cosine transform (DCT). Although this model tracks most transients in the residual signal, it fails at tracking transients which occur near to the end of a DCT block. We improve the model by increasing the frequency resolution of the DCT domain signal and are thus able to detect and model transients in audio signals.

Introduction

Joseph Fourier was the first, who stated with his theorem, that any periodical function (signal) can be modeled as a sum of sinusoids at various amplitudes and harmonic frequencies. Based on this assumption the first sinusoidal model for speech coding presented by McAulay and Quatieri [1], considers that the signal (sound) consists only of periodic components (deterministic part of a signal). The signal is then modeled as a set of sinusoids with time varying frequency, amplitude and phase.

Smith and Serra [2] extended the SM model with non-periodic components (Stochastic part of the signal). The stochastic part of the signal is obtained by filtering white noise with the spectral envelope of the difference signal between original signal and synthesized signal with the SM model (see Fig. 1: switcher in A). This technique is called spectral modeling synthesis (SMS). Although this model improved the SM one by modeling noise with an appropriate technique, it still does not take into account that the stochastic part of the signal is composed of two components: transients and noise (see Fig. 1: switcher in B).

Modeling transients with SM model is not efficient since transients are not slowly varying sine waves. Therefore, their Short-time Fourier transform (STFT) would not contain meaningful peaks. Transients will not also fit well into the noise model, because of the loss of the sharpness in their attack. T.S.Verma and T.H.Y.Meng have proposed an improvement of the SMS model with a transient modeling synthesis approach (TMS) [3]. In this paper, we will present the slightly modified TMS model. In Section 2, we will introduce the TMS technique and extend it with our improvement, Section 3 will show results of the application to percussion music. We will finally close with a conclusion and outlook.

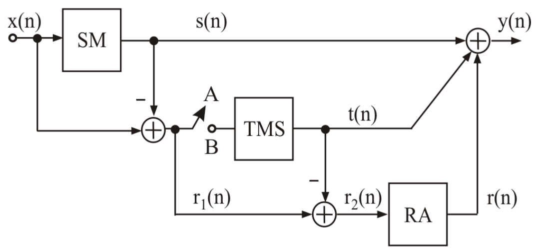

Figure 1: *Sinusoids + Transients + Noise Model. SM (Sinusoidal Modeling), TMS (Transient Modeling Synthesis), RA (Residual approximation). s(n): sinusoids, t(n): transients, r(n): noise*

Transient Modeling Synthesis

The spectrogram of a signal composed of all three components (sinusoids, transients and noise) will be characterised by the presence of horizontal lines for sinusoids and vertical lines for transients (see Fig. 3). Since the horizontal lines are exactly what the SM technique models accurately, we thus need to map the vertical lines into horizontal lines in order to detect and model transients by performing sinusoidal modeling in a properly chosen frequency domain.

Parameters which will characterised the so obtained sinusoids will also characterise the transients components in the time domain signal. Slowly varying sinusoidal signal looks impulsive in the frequency domain, owing to the duality between time and frequency, transients which are impulsive in time domain must then look sinusoidal in the frequency domain. The TMS model as proposed by T.S.Verma and T.H.Y.Meng [3] is based on this duality.

In the analysis part, the first step of the TMS model will be to map the time domain transient signals into sinusoidal signals in some frequency domain. A transformation which can provide this mapping is the discrete cosine transform (DCT). Transients at the beginning of a DCT block will be mapped into low frequency DCT-domain signal, while transients at the end of a DCT block will be mapped into high frequency DCT-domain signal (see Fig. 2). Performing sinusoidal modeling on these DCT-domain signals will correspond to time-domain transients modeling. In order to accurately detect transients with the TMS model, a DCT block length corresponding to about 1 sec (30-60 Sinusoidal modeling frames) of an audio signal is sufficient. Nonoverlapping DCT block should be applied to avoid the modification of transients position in the synthesized signal (see Fig. 2).

In the synthesis part, we first perform sinusoidal synthesis followed by inverse discrete cosine transform (IDCT). Since the TMS model is block based, we may repeat the same procedure up to the last DCT block. Due to the fact that the DCT is applied to nonoverlapping blocks, the whole signal will be reconstructed by concatenating the results after IDCT.

Improved transient model

Although the TMS model detects most transients in the residual signal, we found out that it was not able to detect transients which occur toward the end of a DCT block. This failure of the TMS model is due to the loss of frequency resolution of the DCT-domain signal in that area. We may also remind, that the TMS applied to a DCT blocklength corresponding to round about 10 sinusoidal modeling frames, is only able to locate transient areas in a signal. Since "zero-padding" in one domain results in an increased sampling rate in the other domain, we have then improved the model by zero-padding every signal block before performing DCT on them. We finally find out, that zero-padding with half the DCT block length is sufficient enough to locate and well model transients in audio signals.

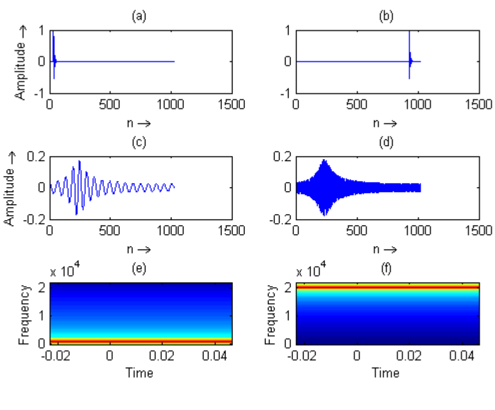

Figure 2: *Transient at the beginning of DCT block (a); transient at the end of DCT block (b); low frequency DCT domain signal (c); high frequency DCT domain signal (d); low frequency sinusoid (e); high frequency sinusoid (f).*

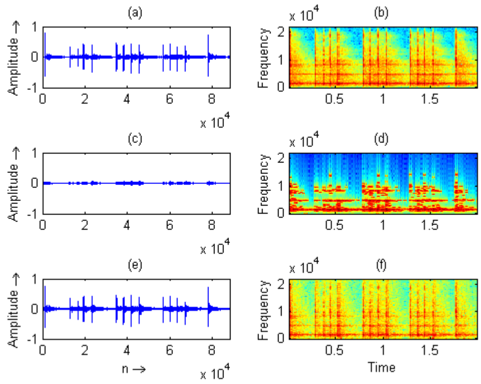

Figure 3: *Original castanet signal (a); spectrogram of original signal (b); SM result (c); spectrogram of SM result (d); STN result (e); spectrogram of STN result (f).*

Application to Music signals

We have applied the modified TMS model to several percussion music signals and are able to detect all the transients. Best results are achieved when we first detect and model transients from the original signal and then perform sinusoidal modeling on the first residual. In Fig. 3 we show the original castanet signal, the SM synthesized signal, the STN synthesized and their respective spectrogram.

Conclusion

The TMS approach as presented in this paper has proved to give good results when applied to audio signals in general and percussion music in particular. We have shown that performing sinusoidal modeling on the DCT-domain signal corresponds to time-domain transients modeling. With the combination of DCT and SM we are able to find meaningful peaks for transients synthesis. Performing sinusoidal modeling on the entire signal may sometime give good results, but it is not recommended since transients and noise do not fit well with that model. The three components model (STN) should be applied on the entire signal for better reconstruction. In future work, we will focus on a transient modeling approach based on envelope detection.

References

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

[2] Xavier Sera 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.

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