A TRANSIENTS / SINUSOIDS / RESIDUAL APPROACH FOR AUDIO SIGNAL DECOMPOSITION

Francois Xavier Nsabimana and Udo Zölzer

Helmut-Schmidt-University / University of the Federal Armed Forces Hamburg, Germany
Department of Signal Processing and Communications
fransa, udo.zoelzer@hsu-hh.de

ABSTRACT

In this paper, we propose an algorithm which splits an audio signal into transient, sinusoidal and residual components. Our algorithm tracks first the transient components in each signal frame and leaves a first residual only composed of sinusoidal and residual components. To achieve this, a transient detection is performed based on the prediction error, its smoothed estimated Hilbert envelope and the first order statistical moments of the latter. Sinusoid detection is then started on the first residual signal using partial tracking taking advantage of psychoacoustics. In the second residual, obtained after subtracting the psychoacoustic relevant sinusoidal components from the first residual, a second partial tracking is started to remove the undesirable sinusoidal components. This ensures a final residual signal free of transient and sinusoidal components. This third residual signal is finally modelled with filtered white noise. Transients are represented with transform coding techniques, while sinusoidal components are modelled as sum of sinusoids with slowly vary amplitude and phase. The synthesized output signal is then obtained by summing up transient, sinusoidal and residual components. This approach has been successful tested with various audio signals. Subjective listening tests of the application to different audio signals are also presented.

1. INTRODUCTION

Originally used in the field of speech data compression, the sinusoidal modeling (SM) approach by McAulay and Quatieri [1] has obtained big attention for audio signal decomposition in computer music. The limitations observed in the SM approach, regarding stochastic signal components, are partially addressed with the spectral modeling synthesis (SMS) approach by Serra [2]. Although the SMS approach addresses the stochastic components of an audio signal, it does not account for Transients. Approaches which account for the transient components are therefore needed. In this paper, we propose an algorithm which splits an audio signal into transient, sinusoidal and residual components (TSR). The remainder of this paper is organized as follows: Section 2 presents our audio signal decomposition approach, Section 3 shows some simulation results and Section 4 concludes.

2. PROPOSED DECOMPOSITION APPROACH

As depicted in Fig. 1 and detailed in [3], we propose here an algorithm which splits an audio signal into transient, sinusoidal and residual components.

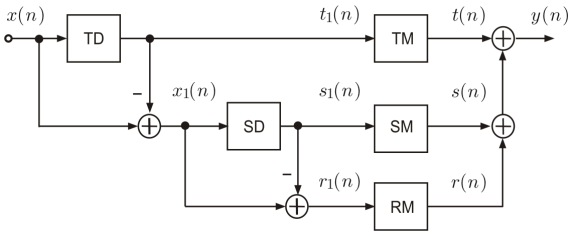

Figure 1: TSR approach. Transient Detection (TD), Transient Modeling (TM), Sinusoid Detection (SD), Sinusoidal Modeling (SM), Residual Modeling (RM). Transient $t(n)$, sinusoid $s(n)$, residual $r(n)$.

2.1. Transient detection approach

In a first step, the TSR approach starts the detection of transients. The main steps of the detection procedure described in [3], are depicted in Fig. 2. The detected transients are then encoded using

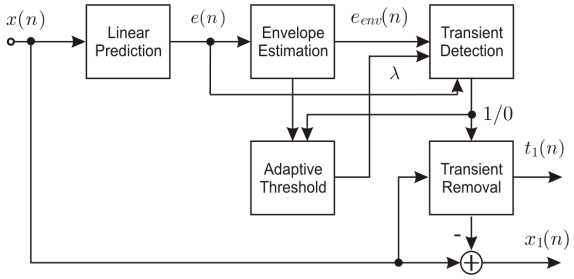

Figure 2: TD: Transient detection and removal approach. Detected transient $t_1(n)$, first residual signal $x_1(n)$.

methods presented in [3].

2.2. Sinusoid detection and modeling

In a second step, sinusoids are detected and modelled using the signal derivative approach in [4]. Based on continuous derivative knowledge, a discrete derivative of the first residual signal

$$x_1^{(1)}(n) = f_s \cdot [x_1(n) - x_1(n-1)], \tag{1}$$

is computed. Here denotes $x_1^{(1)}(n)$ the first signal derivative of $x_1(n)$ and f_s is the sampling rate. After windowing $x_1(n)$ and

$x_1^{(1)}(n)$, DFT is applied and the accurate frequencies of the spectral local maxima are obtained from

$$f_m = \frac{f_s}{\pi} \cdot \arcsin\left(\frac{1}{2f_s}\frac{|X_1^{(1)}(k_m)|}{|X_1(k_m)|}\right). \quad (2)$$

Here denotes $X_1^{(1)}(k_m)$ the DFT value of the 1^{th} signal derivative at local maximum k_m and N is the DFT length. Finally, the corresponding accurate amplitudes a_m and phases φ_m are derived from

$$S_1(f_m) = \sum_{n=0}^{N-1} x_{1w}(n) \cdot e^{-j2\pi\frac{f_m}{f_s}n} = a_m \cdot e^{j\varphi_m}, \quad (3)$$

where $m = 1, \ldots, M$ and M is the number of accurate frequencies. To remove the perceptual irrelevant sinusoids, psychoacoustic masking phenomena are taken into account. In the synthesis stage, the sinusoidal components

$$s(n) = \sum_{m=1}^{M} a_m \cdot \cos\left(2\pi\frac{f_m}{f_s}n - \varphi_m\right) \quad (4)$$

are modelled as sum of sinusoids [1]. Here is $n = 0, \ldots, N-1$, M is the number of retained frequencies and N is the block length.

2.3. Residual modeling

In a third and last step, a residual enhancement (s. Fig. 3) is first performed to ensure a final residual signal free of transient and sinusoidal components. The final residual signal $r(n)$ is then

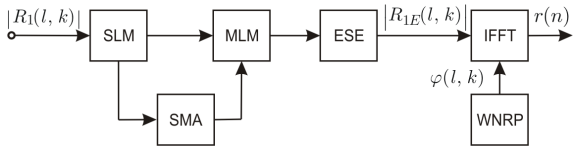

Figure 3: Residual enhancement. Second residual magnitude spectrum $|R_1(l,k)|$, spectral local maxima (SLM), spectral moving average (SMA), modified local maxima (MLM), enhanced spectral envelope (ESE), enhanced magnitude spectrum $|R_{1E}(l,k)|$, white noise random phase (WNRP) and final residual $r(n)$.

modelled as filtered white noise [2].

3. SIMULATION RESULTS

The results with the TSR approach for various signals from the SQAM Database have been evaluated during subjective listening (s. Fig. 5). The subjects had first to find and assign a grade of 100 to the hidden reference signal. They then had to compare the reference signal with results from TSR approach, where transients are encoded using the four methods presented in [3]. Fig. 5 shows that the TSR approach provides very good results.

4. CONCLUSION

We have presented here an audio signal decomposition approach which splits audio signals into transient, sinusoidal and residual components. A residual Enhancement approach prior to residual

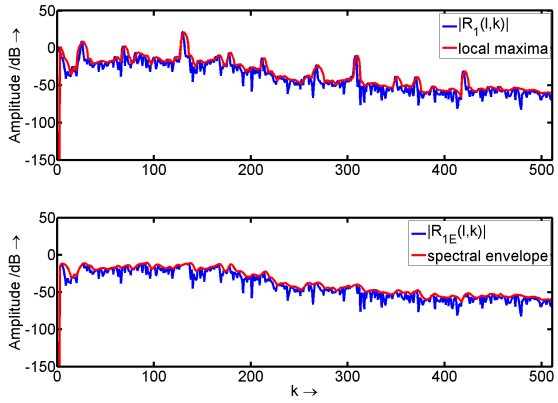

Figure 4: Residual enhancement for the glockenspiel sound file. Second residual magnitude spectrum at frame l $|R_1(l,k)|$ top blue, interpolated local maxima top red. Enhanced magnitude spectrum $|R_{1E}(l,k)|$ bottom blue, obtained spectral envelope bottom red.

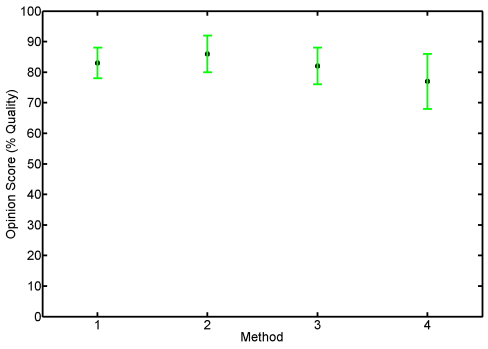

Figure 5: Results from listening test using headphones. Bars denote 95 % confidence interval

modeling has been proposed. An evaluation of the results from the TSR approach has been carried out in form of subjective listening test.

5. REFERENCES

[1] R. McAulay and T. F. Quatieri. Speech analysis/synthesis based on a sinusoidal represenatation. *IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 34, pp. 744-754*, 1986.

[2] 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, vol. 14(4), pp. 14-24*, 1990.

[3] F.X. Nsabimana and U. Zölzer. Transient encoding of audio signals using dyadic approximations. In *Proc. 10th International Conference on Digital Audio Effects (DAFx-07), Bordeaux, France, pp. 297-304*, September 2007.

[4] Sylvain Marchand. *Modélisation informatique du son musical*. PhD thesis, Université de Bordeaux, 2000.