Transient Encoding of Audio Signals Using Dyadic Approximations

Article · October 2002

CITATIONS
5

READS
119

2 authors, including:

Francois Xavier Nsabimana
Helmut Schmidt University
9 PUBLICATIONS **20** CITATIONS

SEE PROFILE

TRANSIENT ENCODING OF AUDIO SIGNALS USING DYADIC APPROXIMATIONS

Francois Xavier Nsabimana and Udo Zölzer

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

ABSTRACT

In this paper, we present a frame based approach for transient detection and encoding of audio signals. The transient detection procedure, as presented here, uses linear prediction within a signal frame followed by an envelope estimation to build an adaptive threshold. Detected transients will automatically be separated and the gaps left by the removed transient are filled with samples from forward and backward extrapolation. To encode detected transients, dyadic approximation approaches are discussed. Results of the application to different audio signals are also presented.

1. INTRODUCTION

An audio signal is generally composed of two parts: a deterministic and a stochastic part. The deterministic part of an audio signal consists of sinusoids, while noise and transients constitute the stochastic part. Although some models as proposed in [1, 2, 3] represent well sinusoids and noise, they really fail for transients. Since transients do not fit well into sinusoids and noise models, they therefore need their own model. In order to build a three components approach (transients + sinusoids + noise) of an audio signal, we need to split the stochastic part into two parts.

In [1] a model for sinusoids with time-varying amplitudes, phases and harmonic frequencies has been presented. In [2], the sinusoidal model (SM) proposed in [1] was extended with a noise model based on residual approximation. The Spectral Modelling Synthesis (SMS) presented in [2] 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.

In [4, 5, 6, 7, 8] and many other recent publications, transient representation is investigated. The methods proposed can be classified into three categories: time domain approach, frequency domain approach and hybrid approach [9]. In this paper, we address the problem of a transient representation under the hybrid approach. We first perform detection in time domain based upon linear prediction followed by encoding in frequency domain using dyadic approximation. In our three components approach (Fig. 1), sinusoids and noise are represented using techniques proposed in [2, 10].

In Section 2, we will present our transient detection approach, Section 3 will introduce the transient encoding approaches, Section 4 will show some simulation results, we will finally close with a conclusion and outlook in Section 5.

2. TRANSIENT DETECTION

Since transients are poorly represented by sinusoids or noise model, it is preferable to represent them separately and leave sinusoids and noise to their own models. We propose, in Fig. 1, to first detect and separate transients within a signal frame. Transients represented as filtered noise will certainly lose their sharpness and sound bad. In Fig. 2 the pre-processing steps for the transient detection 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 transient detector should thus reveal the presence of hidden transients and then emphasize them. In order to successfully 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 accentuate transients in the audio signal.

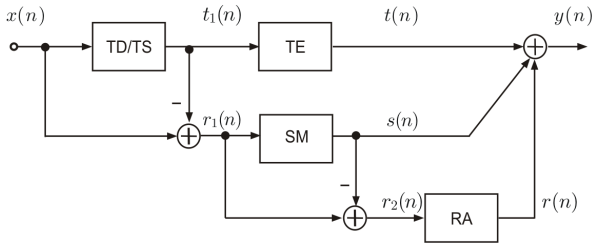

Figure 1: *Transients + Sinusoids + Noise approach. TD/TS (Transient Detection / Transient Separation), TE (Transient Encoding), SM (Sinusoidal Modelling), RA (Residual approximation). s(n): sinusoids, t(n): transients, r(n): noise.*

2.1. Linear Prediction

Assume that signal sample $x(n)$ of an audio signal is to be estimated combining p previous samples. The estimated sample $\hat{x}(n)$ will then be obtained using a finite impulse response (FIR) filter given by

$$\hat{x}(n) = \sum_{i=1}^{p} a_i \cdot x(n-i), \tag{1}$$

where a_i are the filter coefficients obtained by minimizing the square of the prediction error

$$e(n) = x(n) - \hat{x}(n) = x(n) - \sum_{i=1}^{p} a_i \cdot x(n-i) \tag{2}$$

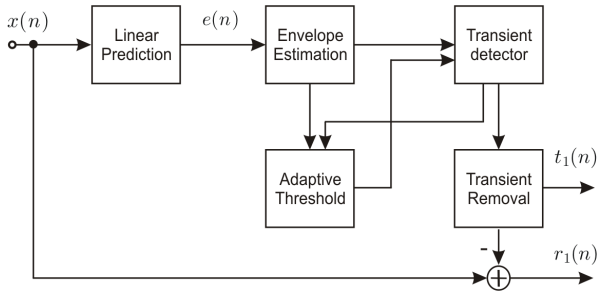

Figure 2: *TD/TS: Transient Detection & Separation approach.* $t_1(n)$*: detected transients,* $r_1(n)$*: first residual.*

within a signal frame. Transforming Eq. (2) into the Z-domain, we obtain the following filter transfer function

$$A(z) = \frac{E(z)}{X(z)} = 1 - P(z) = 1 - \sum_{i=1}^{p} a_i \cdot z^{-i}. \qquad (3)$$

The filter $A(z)$ is designed so that all its zeros are inside the unit circle $|z| = 1$ (excluding the unit circle itself). If we need to recover the filtered input signal $x(n)$ from the error signal $e(n)$, we then have to apply the inverse (synthesis) filter

$$H(z) = \frac{Y(z)}{E(z)} = \frac{1}{1 - P(z)} = \frac{1}{A(z)}. \qquad (4)$$

on $e(n)$.

2.2. Envelope estimation

The simple way to estimate the temporal envelope of a signal $x(n)$ is to take the absolute value of $x(n)$ and apply smoothing using low-pass filter or peak detector. In our model (see Fig. 2), we choose a common and very efficient technique based on the Hilbert Transform (a 90 degree phase shifter) to estimate the envelope of the prediction error signal. With that technique the envelope of a signal $x(n)$ is computed using the corresponding analytic signal

$$\tilde{x}(n) = x(n) + j \cdot \hat{x}(n), \qquad (5)$$

where $\hat{x}(n)$ is the Hilbert transform of $x(n)$. The envelope of the original signal is then simply the modulus of the analytic signal given by

$$x_{env}(n) = |\tilde{x}(n)| = \sqrt{x^2(n) + \hat{x}^2(n)}. \qquad (6)$$

Finally, the envelope is smoothed using a first order low-pass filter.

2.3. The proposed method

The main steps for transient detection as depicted 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 $p = 8$ 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 error 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. The weighting factor is obtained by dividing the maximum value with the mean value of the envelope in the current 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 those indexes where a transient is occurring in the current frame (see Fig. 3 - 4 lower right). In order to avoid multiple detections of the same transient in the overlapping zone, a strategy has been developed. That is if a transient is fully embedded in a frame, a detection flag is triggered. If a fully embedded transient spans over both half-frame, the index of the detected transient corresponding to the second half-frame are used in the next frame for comparison. If a transient spans over successive frames, a detection flag is not triggered since the transient data over these frames needs to be assembled. The gaps left when detected transients are removed, are filled with samples using forward and backward extrapolation as presented in [11]. A three components approach is built, sinusoids and noise are thus left to their respective models, while transients are encoded separately.

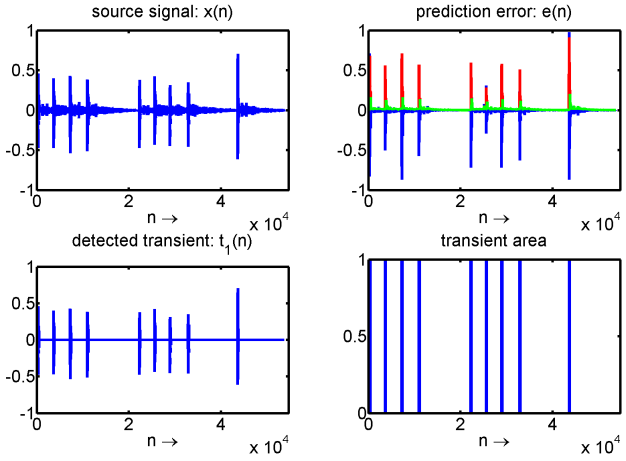

Figure 3: *Transients detection in Castanets sound file.*

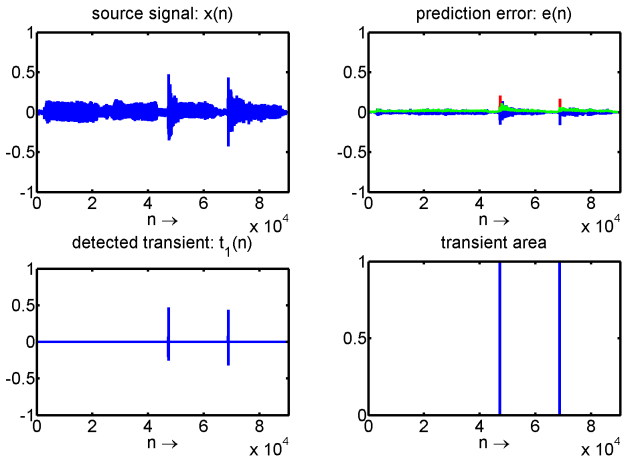

Figure 4: *Transients detection in ABBA sound file.*

2.4. Signal extrapolation

In [11, 12] extrapolation and signal restoration of damaged or removed samples have been deeply investigated. The underlying idea is here the same like in linear prediction, with the only difference that extrapolation needs forward and backward predictor. In our approach the samples to be extrapolated are those corresponding to the removed transients samples (N_2). The procedure for the extrapolation of the missing samples is depicted in Fig. 5, while the results are shown in Fig. 6. The main steps of the extrapolation procedure can be explained as follow:

- Determine the number N_2 of missing samples x_2 within a frame.
- Compare the number N_1 and N_3 of known signal samples (x_1 and x_3) with the number N_2 of missing samples x_2.
- For forward extrapolation: if there are fewer known signal samples than missing samples ($N_1 < N_2$), take known samples from previous frame.
- For backward extrapolation: if there are fewer known signal samples than missing samples ($N_3 < N_2$), take known samples from next frame. In this case backward extrapolation is done in the next frame.
- Apply autoregressive (AR) model of order $p \leq N_2$ to calculate the filter coefficients: a_{if} and a_{ib}.
- Initialize the filter with p past known samples just before the section to be extrapolated.
- Generate a vector of zeros with length N_2, feed it together with z_{if} or z_{ib} as input to the extrapolation filters.
- The output of the two filters are the N_2 extrapolated samples (x_{ef} and x_{eb}).
- Finally sum the forward extrapolated samples and backward extrapolated samples weighted with an appropriate window function (see Fig. 6).

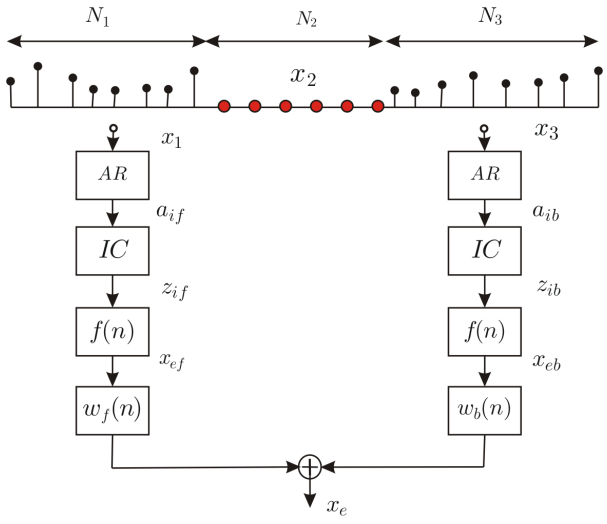

Figure 5: *Extrapolation procedure: autoregressive parameter estimation (AR), Initial Conditions for filter implementation (IC), filter (f(n)), appropriate window function ($w_f(n), w_b(n)$).*

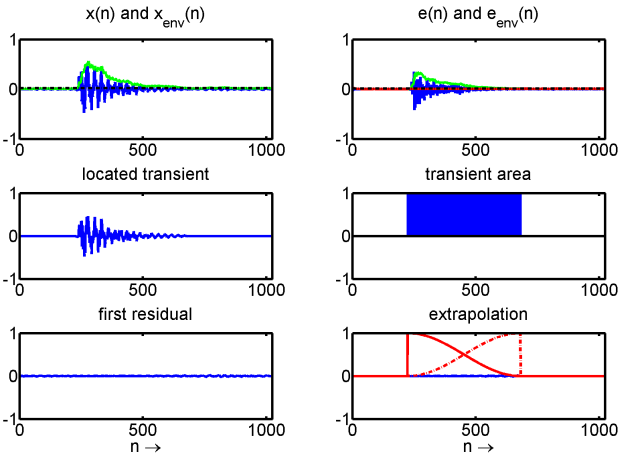

Figure 6: *Transients detection in Castanets sound file.*

3. TRANSIENT ENCODING

We have already stated that transients do not fit well into sinusoids and noise models. Since transients need a very good time resolution, while sinusoids require a very good frequency resolution, a transformation which works with variable resolution is therefore needed. Methods based on dyadic approximation such us multiscale approximation or octave splitting, as presented in Fig. 7, address the problem of multiresolution. An important dyadic approximation method, which has gained increasing attention during the last years, is the discrete wavelet transform (DWT). In [5, 6] transients detection and encoding based on wavelet transform and Hidden Markov tree are presented, while in [13, 14] methods based on iterated filter bank are used. Let $|X(f)|$ be the magnitude spectrum of a signal $x(n)$, using dyadic approximation, the spectrum of $x(n)$ can be first split into two equal parts: low-pass band and high-pass band. These two bands can again be decomposed into

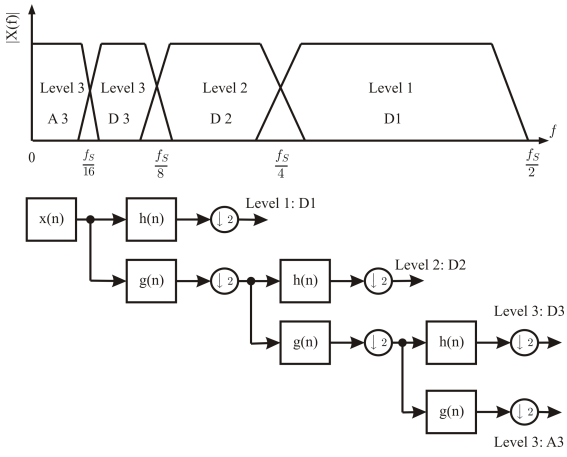

Figure 7: *DWT in frequency domain upper, a 3 level filter bank lower.*

subbands until we reach the number of bands needed for the application (see Fig. 7). Since discrete wavelet transform (DWT) is an implementation of wavelet transforms as an iterated filter bank,

the LP and HP, within one level (see Fig. 7) will represent respectively the scaling filter and the wavelet filter. The signal $x(n)$ is simultaneously decomposed using low-pass filter g and high-pass filter h (see Fig. 7 lower) yielding approximation coefficients

$$A_j(k) = \sum_{i=1}^{N_x+N_F-1} x(i) \cdot g(2k-i) = (x * g)(n) \downarrow 2, \quad (7)$$

and detail coefficients

$$D_j(k) = \sum_{i=1}^{N_x+N_F-1} x(i) \cdot h(2k-i) = (x * h)(n) \downarrow 2, \quad (8)$$

where j is the index of the considered subband, while N_x and N_F are respectively the length of the input signal $x(n)$ and the length of the impulse response g or h (see Fig. 8). The filters g and h are related to each other and must satisfy the quadrature mirror filter relationship.

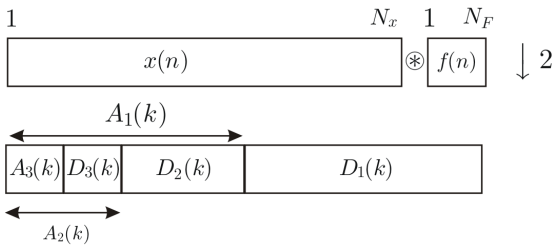

Figure 8: *A 3 level Wavelet decomposition, decomposition filter f(n) (g(n) , h(n)), input signal x(n), detail coefficients ($D_j(k)$) and approximation coefficients ($A_j(k)$).*

In this section, we present our transient encoding approach which performs quantization of the prediction error from discrete cosine transform (DCT) applied on DWT coefficients. We will compare various combinations of dyadic approximations applied to the detected transient. The first method will combine discrete wavelet transform (DWT) and coefficients thresholding within each band. The second method is our proposed approach. The third method is a modified version of the second method, where the discrete cosine transform is not applied. The last method deals with quantization of the prediction error from DCT transformed transient signal. Regarding fame, minimum of smoothness and reduced number of coefficients, we have chosen Daubechies wavelets with 4 vanishing moments (zero moments) to decompose and reconstruct the transients.

3.1. Method I - DWT / Thresholding

This method, as depicted in Fig. 9, deals with discrete wavelet transform (DWT), coefficient thresholding within a subband and inverse discrete wavelet transform (IDWT) for signal reconstruction. The aim of this method is to investigate how far we can reduce the number of coefficients in each subband, but still be able to reconstruct the decomposed signal with the few retained coefficients after thresholding. We have seen from the decomposition explained with Fig. 8, that the approximation coefficients are again split into two parts yielding new detail coefficients and new approximation coefficients of the corresponding subband. Regarding importance of the approximation coefficients in the last subband, we decide not to threshold $A_3(k)$. In the analysis part, a detected transient will first be discrete wavelet transformed using Daubechies wavelets (db4: here 4 is for the order or the vanishing moments) with 3 levels. Applying thresholding (Thresh) based on Root Mean Square (RMS)

$$RMS = \sqrt{\frac{1}{N}\sum_{k=1}^{N} D_j^2(k)}, \quad (9)$$

in each band, except the low-pass band (A_3) (see Fig. 7), we have considerably reduced the number of coefficients to be used for reconstruction. To threshold the wavelets coefficients we have com-

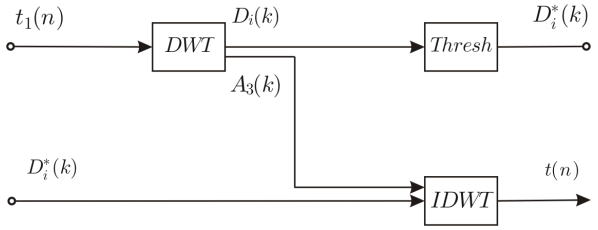

Figure 9: *Method I: Discrete Wavelet Transform (DWT), Thresholding (Thresh) and Inverse Discrete Wavelet Transform (IDWT).*

pared three functions for wavelets coefficients thresholding. The hard-thresholding

$$f_H(D_j) = \begin{cases} D_j & \text{if } |D_j| \geq \lambda, \\ 0 & \text{elsewhere.} \end{cases} \quad (10)$$

retains all coefficients that are greater than the chosen threshold value λ (RMS) and sets to zero others. The soft-thresholding

$$f_S(D_j) = \begin{cases} D_j - \lambda & \text{if } D_j \geq \lambda, \\ 0 & \text{if } |D_j| < \lambda \\ D_j + \lambda & \text{if } D_j \leq -\lambda \end{cases} \quad (11)$$

is shrinkage function, since it shrinks the coefficients by λ towards zero. From Fig. 10 we can see that the hard-thresholding (blue curve) is discontinuous at $|D_j| = \lambda$, while soft-thresholding (green curve) is continuous at $|D_j| = \lambda$ but modifies the value of the retained coefficients. In [15] a custom thresholding function

$$f_C(x) = \begin{cases} D_j - sgn(D_j)(1-\alpha)\lambda & \text{if } |D_j| \geq \lambda, \\ 0 & \text{if } |D_j| \leq \gamma \\ \alpha\lambda(\frac{|D_j|-\gamma}{\lambda-\gamma})^2(\alpha-3)(\frac{|D_j|-\gamma}{\lambda-\gamma}) + 4 - \alpha & \text{elsewhere} \end{cases} \quad (12)$$

where $0 < \gamma < \lambda$ and $0 \leq \alpha \leq 1$ is proposed. This function is a linear combination of hard-thresholding and soft-thresholding, since it combines the advantages of both functions. In Fig. 10, we can recognize that custom-thresholding is equivalent to hard-thresholding with smooth transition around the threshold. From Eq. (12) we can easily go back to Eq. (11) by taking $\alpha = 0$ and to Eq. (10) by taking $\alpha = 1$ and $\gamma = \frac{\lambda}{2}$. For this method, we have finally applied the custom-thresholding function.
The decomposition in subbands is shown in Fig. 9. Except the lower band (A3), the thresholding concerns here only the upper bands. The results of the reconstruction are shown in Fig. 16.

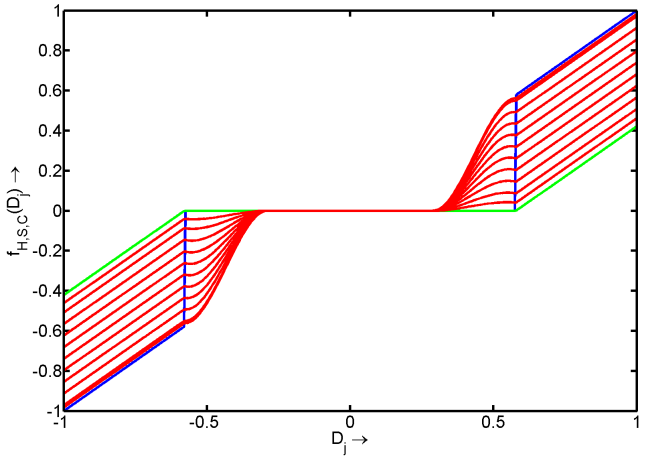

Figure 10: *Coefficients Thresholding: Hard-Thresholding (Blue), Soft-Thresholding (Green) and Custom-Thresholding (Red).*

3.2. Method II - DWT / DCT / LPC / Q

In this method, as depicted in Fig. 11, we present our transient encoding approach. To motivate this choice, we compare this method with various combinations of dyadic approximations applied to the detected transient. The discrete wavelet transform (DWT) followed respectively by discrete cosine transform (DCT), linear prediction coding (LPC) and quantization (Q) are applied on the detected transient. In the analysis part (see Fig. 11 upper), the

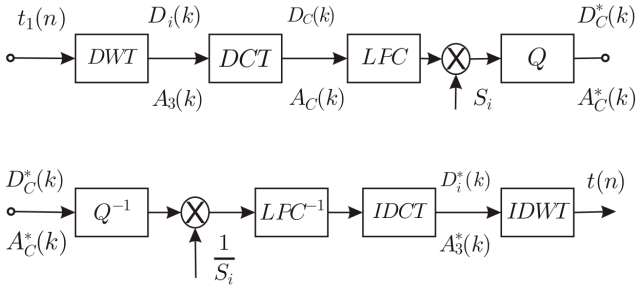

Figure 11: *Method II: Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT), Linear Prediction Coding (LPC) and Quantization (Q).*

DWT coefficients are transformed with the discrete cosine transform (DCT-II) yielding

$$D_{C_j}(k) = \beta(k) \sum_{n=1}^{N} D_j(n) \cos(\frac{(2n+1)k\pi}{2N}) \qquad (13)$$

for the detail coefficients and

$$A_{C_j}(k) = \beta(k) \sum_{n=1}^{N} A_j(n) \cos(\frac{(2n+1)k\pi}{2N}) \qquad (14)$$

for the approximation coefficients, with

$$\beta(k) = \begin{cases} \sqrt{\frac{1}{N}} & \text{if } k = 1, \\ \sqrt{\frac{2}{N}} & \text{for } k = 2, .., N. \end{cases} \qquad (15)$$

where N is the length of the coefficient array in the considered subband and j is the index of the corresponding subband. On the DCT transformed coefficients linear prediction coding (LPC) is then performed. A prediction filter with model order $p = 8$ is used for this method. In each subband the prediction error is first normalized (S_i) before quantization is applied on it. The original 16 bit quantization word-length is here reduced to only 4 bit. In the synthesis stage, the inverse of the scaling factor ($1/S_i$) is first multiplied with the quantized signal before LPC synthesis filter is applied on it. Assuming that the analysis filter $A(z)$ used in the LPC part is the one designed with Eq. (3), we can expect perfect reconstruction of the coefficients using the inverse filter $H(z)$ (Eq. (4)). The inverse discrete cosine transform (IDCT) and inverse discrete wavelet transform (IDWT) are performed for final reconstruction.

3.3. Method III - DWT / LPC / Q

This method, as presented in Fig. 12, is a modified version of the previous method. In the analysis part, linear prediction coding (LPC) is directly applied to the coefficients from the discrete wavelet transform (DWT). We finally quantize the prediction error using the same word-length like in method II. A prediction filter with model order $p = 8$ is again used here. The prediction error is normalized in each subband before quantization is applied. The normalized prediction error is then 4 bit quantized. In the synthesis stage, the inverse of the scaling factor ($1/S_i$) is multiplied with the quantized signal before LPC synthesis filter is applied on it. We finally reconstruct the signal the inverse discrete wavelet transform (IDWT).

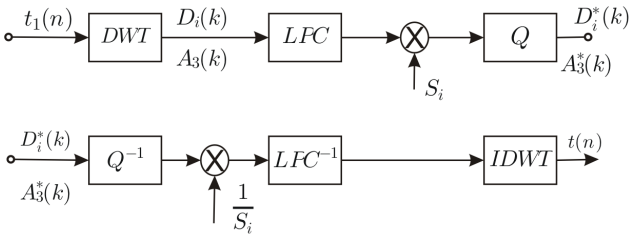

Figure 12: *Method III: Discrete Wavelet Transform (DWT), Linear Prediction Coding (LPC) and Quantization (Q).*

3.4. Method IV - DCT / LPC / Q

The last method, as shown in Fig. 13, deals with the discrete cosine transform (DCT) directly applied to the detected transient signal. The DCT transformed signal is then linear predicted yielding a prediction error which will be finally quantized. The prediction error is first normalized in each subband before quantization is applied on it. The original 16 bit quantization word-length is again here reduced to only 4 bit. In the synthesis stage, the inverse of the scaling factor is first multiplied with the 4 bit quantized signal before LPC synthesis filter is applied on it. We finally reconstruct the signal using the inverse discrete cosine transform (IDCT).

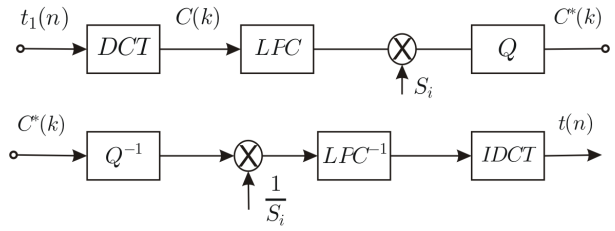

Figure 13: *Method IV: Discrete cosine Transform (DCT), Linear Prediction Coding (LPC) and Quantization (Q).*

4. SIMULATION RESULTS

We have applied the four methods to detected transient signal and have calculated the L2-norm

$$l = \sqrt{\sum_{n=1}^{N} r^2(n)} \tag{16}$$

of the residual signal from each method. In [16], a similar approach has been used for comparison of reconstruction results in image processing. We should point out that this way of characterizing the residual signal is purely numerical and does not take perceptual hearing considerations into account. Although there might be a correlation between numerical error and sound quality. In Fig. 15 results obtained when applying only the Spectral modelling Synthesis (SMS) to the castanet signal are shown. The original signal $x(n)$, the reconstructed signal $y(n)$ and the residual signal $r_{x-y}(n)$ with the L2-norm are presented. From Fig. 16 up to Fig. 19, results obtained with approaches depicted in Fig. 1 applied to the same castanet samples are shown. We can easily recognize that the reconstructed signal $y(n)$ is very close the original signal $x(n)$ for all the methods, with small differences regarding their L2-norm. We can also notice that the three components approach really outperforms the Spectral modelling Synthesis (SMS) approach for this kind of signal. We have applied the same simulation to different audio signals. For the sake of completeness, we will show here the results obtained with glockenspiel samples. Fig. 20 shows results obtained when applying only the Spectral modelling Synthesis (SMS) to the glockenspiel signal. In Fig. 21 results obtained when applying the SMS approach combined with the second transient encoding approach to the glockenspiel samples are presented. It is worth mentioning that the three components approach again outperforms the Spectral modelling Synthesis (SMS) approach regarding L2-norm. Similar results are also obtained applying the same approaches to samples from ABBA sound file (see Fig. 22). Improvement of the SMS approach is expected with the new promising method proposed [17]. This method analyzes the noise without any prior knowledge of the sinusoids model as presented in [2]. For all the methods, where linear prediction coding (LPC) is used, model order $p = 8$ has been applied. The original 16 bit quantization word-length of the input signal is reduced to only 4 bit for these methods. Regarding L2-norm results, we can notice that the three first methods remain close to each other. The L2-norm of residual signal obtained with method IV is bigger than the one obtained with other three methods. In Table 1 and Fig. 14, results obtained during listening test with headphones are presented. The subjects recruited for this test are all working in our lab. All the subjects assigned a grade of 100 to the hidden reference signal. While Subject F is the only one who graded the proposed method lower than the other methods, subject I even graded method III higher than others. Subject C did not perceive any difference between all the methods. But nevertheless method II is graded the best and method IV is scored lowest by all the subjects. With results presented in Table 1 and the L2-norm of the residual signal, a correlation between numerical error and sound quality is somehow observed.

Subject	Method.I	Method.II	Method.III	Method.IV
A	80	80	80	40
B	70	80	70	80
C	90	90	90	90
D	75	90	80	60
E	90	93	90	90
F	80	60	60	80
G	78	88	80	75
H	90	90	80	80
I	85	95	98	80
J	92	90	85	90
K	85	90	80	80
L	75	90	85	80
μ	82.5	86.33	81.5	77.08
σ	6.91	9.02	9.34	13.61
Interval	$\pm$ 4.59	$\pm$ 5.99	$\pm$ 6.20	$\pm$ 9.03

Table 1: Results from listening test using headphones. μ is the arithmetic average and σ is the standard deviation. **Interval**: 95 % confidence interval.

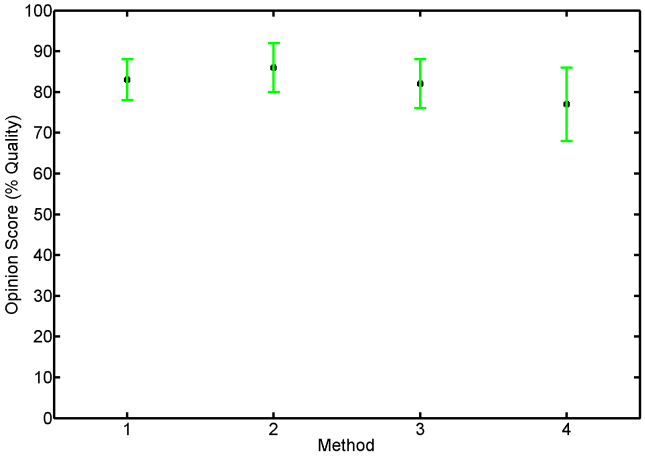

Figure 14: *Results from listening test using headphones. Bars denote 95 % confidence interval*

5. CONCLUSION

We have presented a method for transient detection based upon linear prediction combined with envelope estimation. This method succeeds in detecting transients in various kinds of audio signals. We have shown several alternatives for transient encoding using dyadic approximations. Regarding the L2-Norm of the residual signal, the method based on the discrete wavelet transform (DWT)

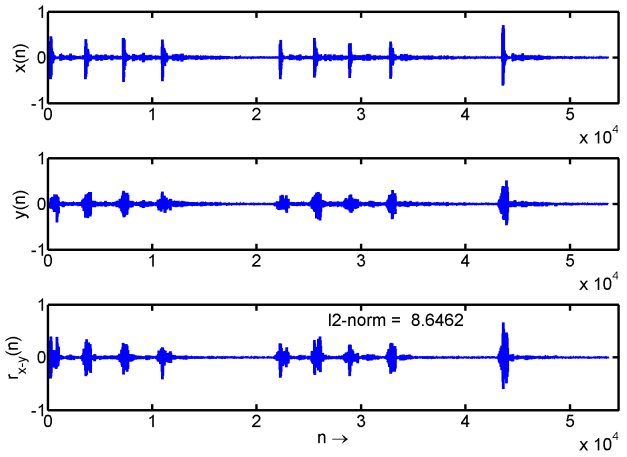

Figure 15: *Original signal x(n) (Castanets), SMS synthesized signal y(n), residual signal* $r_{x-y}(n)$.

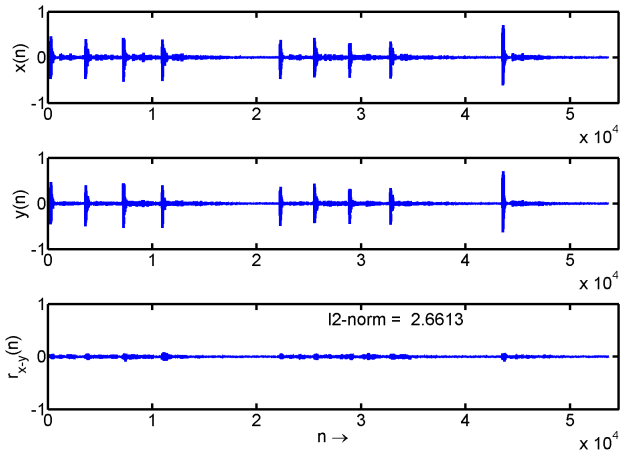

Figure 16: *Original signal x(n) (Castanets), SMS + Method I y(n), residual signal* $r_{x-y}(n)$.

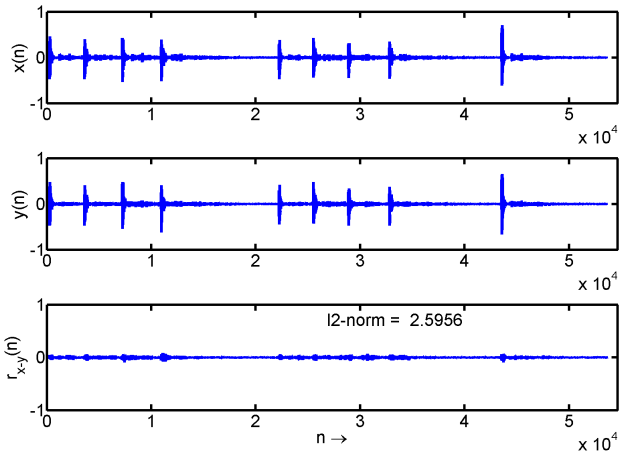

Figure 17: *Original signal x(n) (Castanets), SMS + Method II y(n), residual signal* $r_{x-y}(n)$.

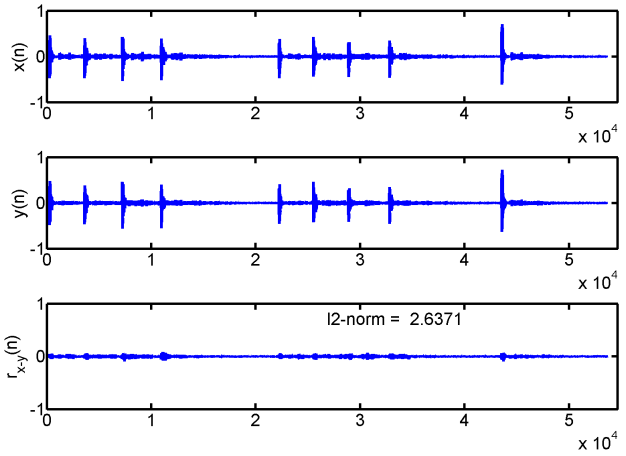

Figure 18: *Original signal x(n) (Castanets), SMS + Method III y(n), residual signal* $r_{x-y}(n)$.

followed by the discrete cosine transform (DCT), Linear Prediction Coding (LPC) and 4 bit quantization remains close to method I and III for all the tested signals. Subjective listening tests combined with the L2-Norm of the residual signal, indicate exactly that method II outperforms others. For future work, the same listening tests will be repeated in modified order and with many different audio files to observe if the trend remains the same.

6. 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] M. Desainte-Catherine and S. Marchand. High precision fourier analysis of sounds using signal derivatives. In *J. Audio Eng. Soc. ,vol. 48 N° 7/8, pp. 654-667*, May 2000.

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

[5] S. Molla L. Daudet and B. Torrésani. Transient detection and encoding using wavelet coefficient trees. In *Proc. of the GRETSI'01 conference, F. Flandrin Ed.*, 2001.

[6] Molla and B. Torrésani. Hidden markov tree based transient estimation for audio coding. In *Proc. of the conference ICME´02 conference, P.Vandergheynst Ed.*, 2002.

[7] M. Sandler C. Duxbury and M. Davies. A hybrid approach to music note onset detection. In *Proc. 5th International Conference on Digital Audio Effects (DAFx-02), Hamburg, Germany, pp. 33-38*, September 2002.

[8] J. Uscher. Extraction and removal of percussive sounds from musical recordings. In *Proc. 9th International Conference*

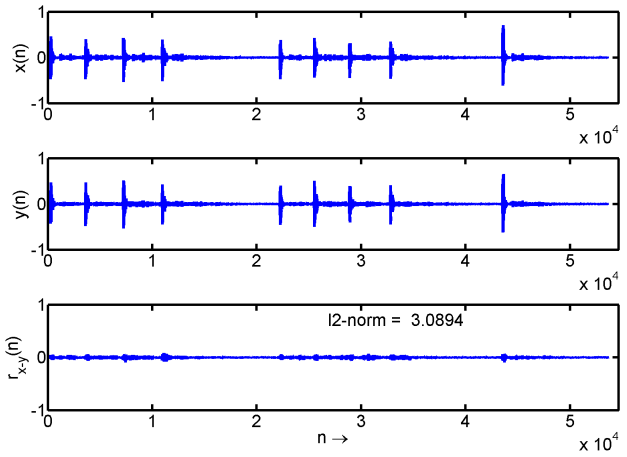

Figure 19: *Original signal x(n) (Castanets), SMS + Method IV y(n), residual signal* $r_{x-y}(n)$.

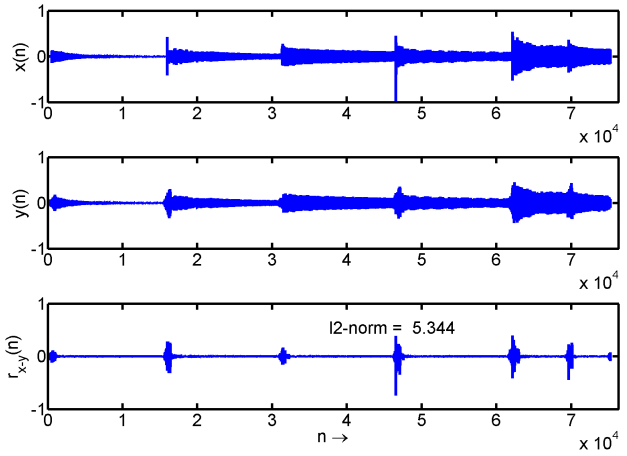

Figure 20: *Original signal x(n) (Glockenspiel), SMS synthesized signal y(n), residual signal* $r_{x-y}(n)$.

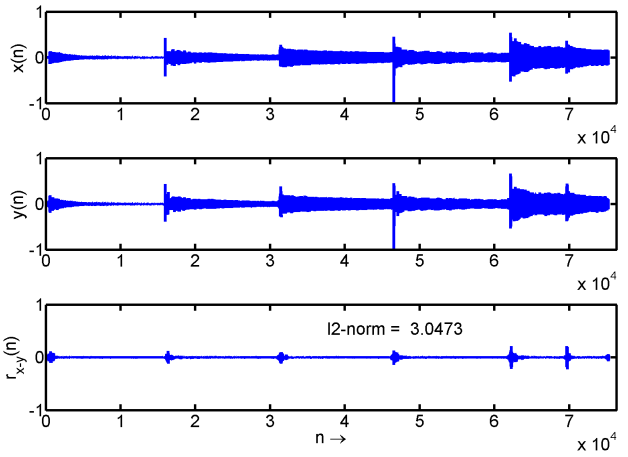

Figure 21: *Original signal x(n) (Glockenspiel), SMS + Method II y(n), residual signal* $r_{x-y}(n)$.

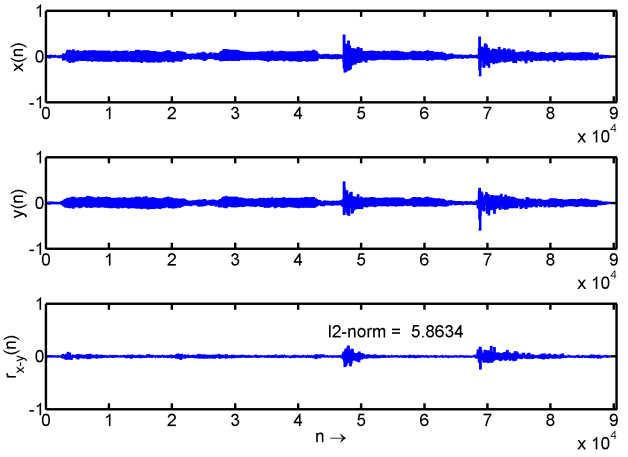

Figure 22: *Original signal x(n) (ABBA sound file), SMS + Method II y(n), residual signal* $r_{x-y}(n)$.

on Digital Audio Effects (DAFx-06), Montreal, Canada, pp. 263-266, September 2006.

[9] L. Daudet. A review on techniques for the extraction of transients in musical signals. In *Proc. of the CMMR´05 conference, Pisa, Italy, pp. 219-232*, 2005.

[10] M. Lagrange and S. Marchand. Real-time additive synthesis of sound by taking advantage of psychoacoustics. In *Proc. 4th International Conference on Digital Audio Effects (DAFx-01), Limerick, Ireland, pp. 5*, December 2001.

[11] I. Kauppinen and K. Roth. Audio signal extrapolation - theory and applications. In *Proc. 5th International Conference on Digital Audio Effects (DAFx-02), Hamburg, Germany, pp. 105-110*, September 2002.

[12] S. V. Vaseghi. *Advanced Signal Processing and Digital Noise Reduction*. Willey &. Teubner, Stuttgart, 1996. ISBN 3-519-06451-0.

[13] M. Davies C. Duxbury and M. Sandler. Separation of transient information in musical audio using multiresolution analysis techniques. In *Proc. 4th International Conference on Digital Audio Effects (DAFx-01), Limerick, Ireland, pp. 1*, December 2001.

[14] M. Sandler C. Duxbury, J. P. Bello and M. Davies. A comparison between fixed and multiresolution analysis for onset detectionin musical signals. In *Proc. 7th International Conference on Digital Audio Effects (DAFx-04), Naples, Italy, pp. 207-211*, October 2004.

[15] B. J. Yoon and P. P. Vaidyanathan. Wavelet-based denoising by customized thresholding. In *ICASSP, Montreal, Canada, vol. 2, pp. ii-925-8*, May 2004.

[16] L. Yaroslavsky and Y. Chernobrodov. Dct and dft discrete sinc-interpolation methods for direct fourier tomographic reconstruction. In *ISPA, Rome Italy, pp. 405*, 2003.

[17] P. Hanna G. Meurisse and S. MARCHAND. A new analysis method for sinusoids + noise spectral models. In *Proc. 9th International Conference on Digital Audio Effect (DAFx-06), Montreal, Canada*, September 2006.