

# Transients Detection in the Time-Scale Domain

V. Bruni and D. Vitulano\*

Istituto per le Applicazioni del Calcolo - C.N.R., Viale del Policlinico 137,  
00161 Rome Italy  
{bruni,vitulano}@iac.rm.cnr.it

**Abstract.** In this paper a novel model for transients detection in piecewise stationary signals is presented. A hybrid representation is assumed for the signal and the different behavior of each component (stationary, transient and stochastic) in the time-scale plane is exploited. Experimental results on both shape contours, described by a differential chain code, and audio signals show the generality of the proposed model.

**Keywords:** Transients, chain code, shape analysis, audio signal, wavelets.

## 1 Introduction

A classical problem of signal and image processing is the detection of signal significant components, such as edges and textures for images or transients and harmonic parts for audio signals. A wide literature copes with this problem since it is somewhat difficult to separate, without ambiguity, different parts of the signal due to their joint interactions (masking, occlusion, etc.) [13]. In audio processing [4,5,6,7,8,14], the signal is modeled as the superposition of three distinct components: *transients*, that correspond to the attack of the notes or abrupt changes in the sound; *tonal component*, that characterizes the harmonic parts of the signal, and a *stochastic residual*, that is a stationary random signal with smooth power spectrum [6]. A similar hybrid structure can be recognized in shape analysis through a 1-D representation of their contours. It contains information about the local changes of the shape, that can correspond to both natural variations of the curvature or to the spatial discretization of the shape. The former is a transient information, for examples corners, while the latter is a tonal component, since it describes the global orientation of the shape in the space. In Fig. 1 there is an example that shows a simple shape and its normalized differential chain code[9]. This latter is a locally stationary signal with transients in correspondence to abrupt changes in the global orientation. For real world shapes the stochastic characterization of the residual component can be weaker. An example are cavities of historical buildings in cultural heritage. The degradation can be found in the residual component whenever transients and harmonic parts are extracted. The different features of the three components allows to find a proper expansion basis for each of them, i.e. the one yielding a sparse representation. In particular, the tonal component shows

---

\* This paper has been partially supported by the CNR RSTL ‘Multiscale Analysis for Complicated Shapes Recognition’.

significant peaks in the power spectrum, while wavelets are more suitable for transients representation, since their good time localization properties. In audio processing this kind of approach leads to transient and steady state separation (TSS) techniques [4]. Conventional methods first separate the tonal component by selecting significant coefficients in a Fourier basis (Discrete Cosine Transform, Molecular Discrete Cosine Transform, etc.). Then, transients are detected by retaining high amplitude wavelet coefficients of the remaining signal. Another possibility is to define harmonic atoms and use a matching pursuit algorithm for extracting the ones contained in the analysed function [8,10]. Nonetheless, there can be two kinds of problems: i) the signal components cannot be well separated; ii) a thresholding is often employed for detecting significant coefficients. This entails recursive projection algorithm for refining the result. The aim of this work is to exploit the multi-scale characterization of transients for their good detection and separation from the remaining components of the signal. The detection is not limited to give a smooth approximation of the signal (Fourier or wavelet descriptors [9]) but it aims to reconstruct their singular contribution over the original signal. In other words the aim is to have a polygonal representation of the analyzed signal whose corners are in correspondence to true changes in curvature of the signal. We refer to the hybrid signal model used in audio signal, but we try to detect transients directly from the original signal, without a pre-analysis of the tonal component. This is possible thanks to the modeling of transients as isolated singularities in the original signal. They show an increasing decay along scales which is opposite to the tonal component one. This allows their detection at coarser scales while their contribution at finer scales can be predicted using a proper atomic approximation. This latter accounts for transients evolution along scales and corresponds to a piecewise linear representation of the signal. This characterization allows an almost faithful detection and representation of transients and their separation from the tonal component. Some preliminary experimental results will show the potentialities of the method and performances comparable to the TSS techniques.

## 2 The Proposed Model for Transients Detection

Let the signal  $f(t)$  be represented as the superposition of three components:

$$f(t) = f_{ton}(t) + f_{tran}(t) + f_{res}(t), \quad \forall t. \quad (1)$$

The goal of this section will be to extract  $f_{tran}$  from  $f$ , where  $f_{ton}$  is a locally stationary part, i.e. it is the composition of harmonic atoms;  $f_{tran}$  is a piecewise smooth signal;  $f_{res}$  is a stochastic residual that does not well match with previous classes. The main peculiarity of a wavelet representation is the characterization of singularities through the decay of their coefficients along scale levels [11]. This is known as clustering and persistency property and it is employed in a probabilistic model to discard real transients coefficients. More precisely, a wavelet coefficient can have a transient or a residual state, respectively characterized by a large and small variance distribution. A tree is constructed using the maximum likelihood from coarse to fine scale with the following rule: if the child is a transient, its

parent cannot be a residual [14]. In the following we will use a formal model for describing the evolution of singularities along scales and we will try to use it for transients detection as a direct analysis of the original signal, without a preprocessing for catching the tonal component.

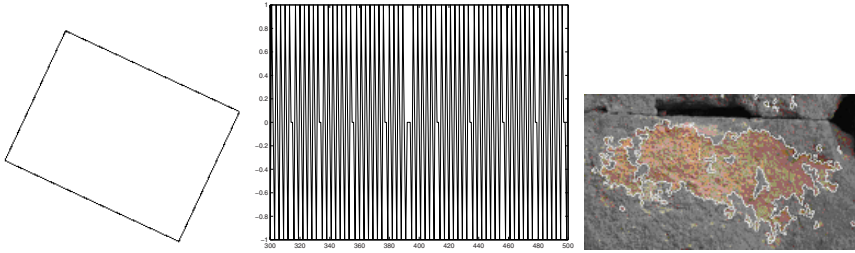
The wavelet theory includes an important result about the characterization of isolated singularities in the time scale. In particular, it is possible to prove that for a signal  $f$  of Lipschitz order  $\gamma$ , the decay of the corresponding wavelet coefficient is proportional to  $s^{\gamma+\frac{1}{2}}$  [12,11], i.e.  $|w(u, s)| \leq As^{\gamma+1/2}$ , where  $w(u, s)$  is the wavelet transform of  $f$  at scale  $s$  and time  $u$  while  $A$  is a constant. In [2], the authors derived the trajectories of modulus maxima corresponding to isolated singularities of order  $\gamma_k$  and they studied their interaction by modeling the coefficients corresponding to a single singularity as waves that travel along scales. These waves are subjected to both a diffusive and source effect till they interfere with other waves. From this point on a transport effect regulates waves interference till they become a single wave. The sourcing effect is weighted by the energy contribution of each wave and it is connected to the order of the singularities. In particular, for the wavelet transform of a piecewise regular signal  $f$  with singularities of order  $\gamma_k$  at  $t_k$ , it is possible to write the following evolution law in the  $(u, s)$  plane:

$$w_s = -\frac{u}{s}w_u + \sum_k \frac{t_k}{s}w_u^{(k)} + \frac{1}{s}w + 2 \sum_k \frac{\gamma_k}{s}w^{(k)}, \tag{2}$$

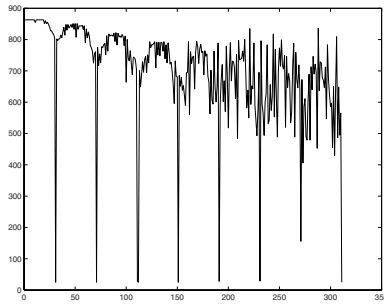
where  $w_s$  and  $w_u$  respectively are the partial derivatives of  $w$  with respect to  $s$  and  $u$ , while  $w^{(k)}$  is the contribution of the singularity located at  $t_k$ . For computational purposes,  $w^{(k)}(u, s) = \alpha_k s^{\gamma_k-1} F(t_k, u, s)$ , where  $F(t_k, u, s)$  is the wavelet transform of a first order singularity located at  $t_k$  of a piecewise linear signal and  $\alpha_k$  is the corresponding slope — see [2] for details. From the previous equation it is possible to derive the trajectories  $u(s)$  of each global maximum of  $w(u, s)$  in the time scale plane, i.e.

$$\begin{cases} \dot{u} = -\frac{t_k-u}{s} - \frac{1}{s} \frac{\sum_h d_{kh} w_{uu}^{(k)} + \gamma_h w_u^{(k)}}{w_{uu}} \\ u(1) = t_k \end{cases}$$

where  $d_{kh} = t_k - t_h$ . If  $2C$  is the support for a symmetric and compactly supported wavelet, a singularity is isolated till its cone of influence ( $|u - t_k| \leq Cs$ ) does not intersect with the one of another singularities. In other words a singularity can be isolated from scale  $s = 1$  till scale  $s = \bar{s}$ . From  $\bar{s}$  on, it interacts with other singularities. From modulus maxima point of view, this movement corresponds to a precise trajectory in the time scale plane: each global maximum is attracted or rejected from the neighboring ones. If it is attracted, it will tend to be confused with the other one, if it is rejected it does not disappear. This kind of interference implies that for a piecewise regular signal the number of global maxima, i.e. the one corresponding to single waves, is constant or decreases along scale levels. On the other hand the tonal component shows a periodic behavior along scale levels, due to its harmonic characterization. It turns out that the



**Fig. 1.** A rectangular shape with a fixed orientation (*left*). Its corresponding normalized differential chain code (*middle*), that is a piecewise stationary signal. A more complicated shape from degradation of Cultural Heritage (*right*).



**Fig. 2.** Number of modulus maxima versus scale of the leftmost shape in Fig. 1

numbers of the corresponding modulus maxima could not vary monotonically, as in Fig. 2. This observation allows to discard scales where the contribution of the tonal component is not negligible with respect to the transient one. They are the scales such that the monotonic non increasing behavior of the number of modulus maxima is interrupted.

Nonetheless a finite number of scales does not allow the inversion of the transform, while dyadic scales may contain a great contribution of the tonal component. The time scale evolution law of wavelet coefficients in eq. (2) can be exploited to predict details at dyadic scales from the ones selected in the previous step, i.e. the ones having a negligible contribution of the tonal component. In fact, the solution of eq. (2) can be written as  $w(u, s) = \sum_k \alpha_k s^{\gamma_k - 1} F(t_k, u, s)$ . This form, called atomic approximation, suggests that the wavelet transform of a generic signal at a fixed scale  $s$  corresponds to the wavelet transform at that scale of a piecewise linear signal whose slopes are defined by equation  $\alpha_{k,s} = \alpha_k s^{\gamma_k - 1}$ , as proved by the following proposition.

**Proposition 1.** *From the atomic approximation at a given scale  $\bar{s}$  of a function  $f$ , it is possible to build a piecewise linear function  $a(t)$  whose wavelet details at the same scale are equal to the ones of a piecewise linear approximation of  $f$ .*

**Proof:** Let us fix the scale level  $\bar{s}$  and let us compute the atomic representation of the wavelet transform at that scale, i.e.  $\{\alpha_{k,s}, t_k\}_{1 \leq k \leq N}$ , hence  $w(u, \bar{s}) = \sum_{k=1}^N \alpha_{k,s} F(t_k, u, \bar{s})$ , with  $\alpha_{k,s} = \alpha_{k,1} s^{\gamma_k - 1}$ . Each atom  $F(t_k, u, \bar{s})$  equals the wavelet transform at scale  $\bar{s}$  of an infinite ramp signal  $r_k(t)$  with a singularity in  $t_k$ , i.e.  $r_k(t) = \begin{cases} 0 & t \leq t_k \\ \alpha_{k,s}(t - t_k) & t > t_k. \end{cases}$  From the linearity of the wavelet transform and indicating with  $W$  the wavelet transform operator, it then follows

$$w(u, \bar{s}) = \sum_{k=1}^N \alpha_{k,s} F(t_k, u, \bar{s}) = \sum_{k=1}^N W r_k(t) = W \left( \sum_{k=1}^N r_k(t) \right) = W a(t),$$

where  $a(t) = \sum_{k=1}^N \left( \left( \sum_{h=1}^k \alpha_{h,s} \right) (t - t_{h-1}) + \beta_h \right) \chi_{[t_{k-1}, t_k]}$ , with  $\beta_h = \sum_{h=1}^{k-1} \alpha_{h,s} (t_k - t_h)$ . ⊙

It is worth noticing that the support signal  $a(t)$  has the same atomic approximation of the signal  $f$  at scale  $\bar{s}$  while its low pass residual is different. In that way, from the atomic approximation of the wavelet transform of a signal at a fixed scale  $\bar{s}$ , it is possible to derive the details at successive scales by simply performing the wavelet transform of the support signal  $a(t)$ . While the slopes  $\alpha_{k,s}$  can be derived using a matching pursuit like algorithm, as described in [3], the decay exponents  $\gamma_k$  can be extracted by comparing two successive scales at atoms locations. In fact, for two different scales  $s_1$  and  $s_2$ , we have  $\alpha_{k,s_1} = \alpha_{k,1} s_1^{\gamma_k - 1}$  and  $\alpha_{k,s_2} = \alpha_{k,1} s_2^{\gamma_k - 1}$ . Hence,  $\alpha_{k,s_2} = \alpha_{k,s_1} \frac{s_2^{\gamma_k - 1}}{s_1^{\gamma_k - 1}}$  and then

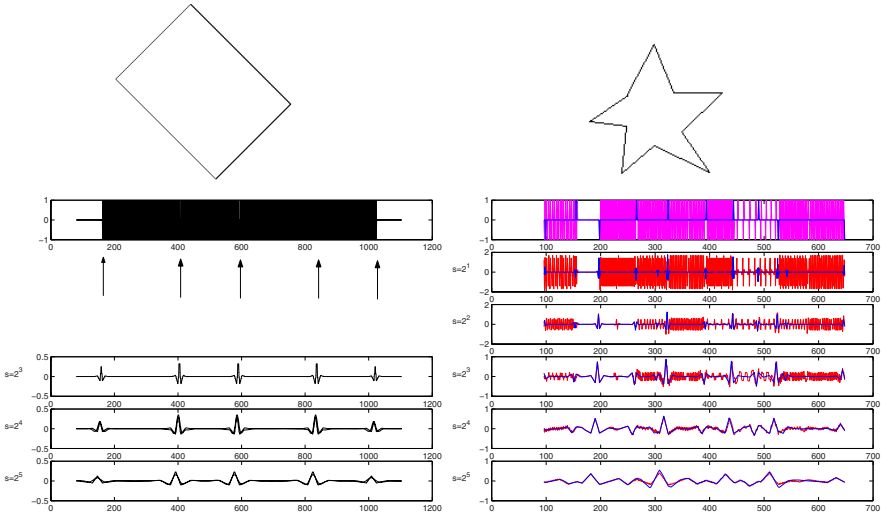
$$\gamma_k = 1 + \frac{\log_2 \left( \frac{\alpha_{k,s_2}}{\alpha_{k,s_1}} \right)}{\log_2 \left( \frac{s_2}{s_1} \right)}. \tag{3}$$

It is worth outlining that the atomic representation preserves the correlation between adjacent coefficients of the wavelet decomposition. This property allows the recovering of transients coefficients under threshold and avoids artifacts due to the rough cut off of information in the thresholding based approaches [3].

In order to avoid the introduction of false transients, the coarsest selected scale can provide the intervals where to find transient contribution at finer scales. In fact, as proved by the evolution law, transient contribution becomes more evident at coarser scales. It turns out that they can be selected with a simple thresholding operation. When all the dyadic scales are reconstructed, the transform can be inverted for getting the transient component  $f_{tran}$ .

### 2.1 The Algorithm

1. Perform the continuous wavelet transform (CWT)  $w(u, s)$  of  $f$  from the scale  $s = 1$  to the scale  $s = S$  using  $p$  as discretization step. Let  $J = \lfloor \log_2(S) + 0.5 \rfloor$
2. Compute the low pass component  $Af(u, 2^J)$  of  $f$  at scale level  $2^J$
3. Sort the number of modulus maxima of  $w(u, s)$  at each scale in increasing order. Let nMAX be the sorted vector



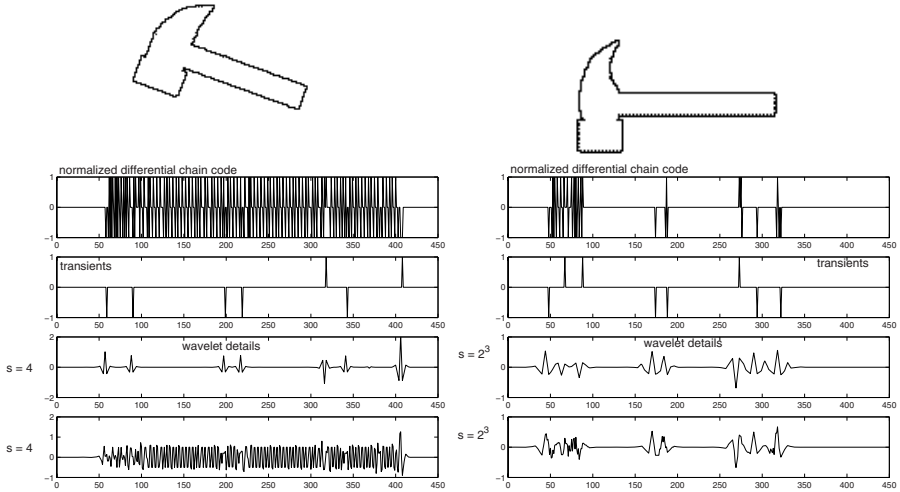
**Fig. 3.** 1<sup>st</sup> column) Rectangle oriented at 45°; normalized chain code and transients locations detected using the proposed scheme; reconstructed transient contribution in the wavelet domain. 2<sup>nd</sup> column) Star shape; the corresponding normalized chain code (magenta) and the estimated  $f_{trans}$  (blue) using the proposed scheme; wavelet coefficients of the transient component (blue) using proper basic atoms.

4. Let  $\hat{s}$  be the scale level having the number of atoms equal to  $nMax(1)$ ; retain its atoms whose amplitude over-exceeds the value  $T = \frac{\|w(u, \hat{s})\|^2}{\hat{s}}$  and select the location of their global maxima  $\{t_{k, \hat{s}}\}$
5. For  $j = 1, 2, \dots, J$ 
  - Select the scale  $\tilde{s}$  having the least number of maxima among the scale levels that satisfy:  $\lfloor \log_2(s) + 0.5 \rfloor = j$
  - Compute the atomic approximation of  $w(u, s)$  estimating the atoms slopes  $\alpha_k$  using the algorithm in [2], and  $\gamma_k$  using  $s_j$  and  $s_j + p$  through eq. (3) in the cone of influence of  $t_{k, J}$  at the selected scale. Built the corresponding support signal  $a(t)$ , as in Prop. 1.
  - Compute the undecimated discrete wavelet transform (UDWT) of  $a(t)$  at scale level  $j$  and let  $w(u, 2^j)$  the achieved detail
6. Invert the UDWT using  $\{w(u, 2^j)\}_{1 \leq j \leq J}$  and  $Af(u, 2^J)$  for getting  $f_{trans}$ .

It is worth noticing that step 5 can be relaxed. In fact, from a scale  $s_j$  we can predict more than one dyadic scale. In this way it is also possible to drastically reduce the computational effort.

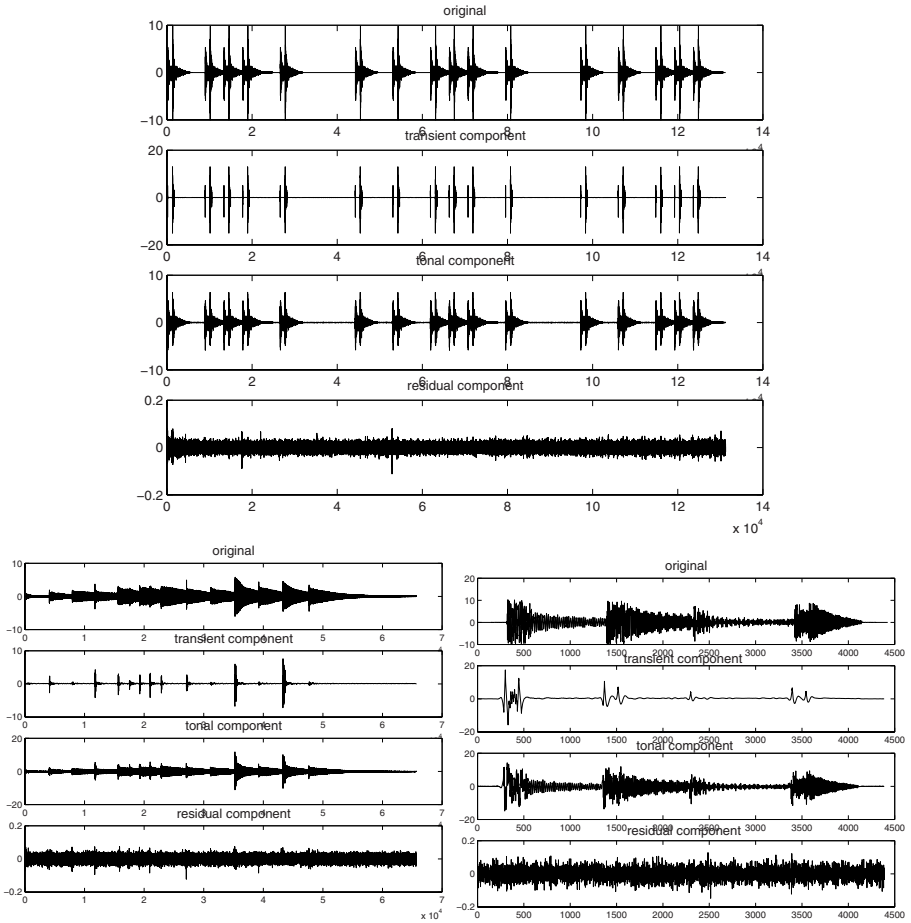
### 3 Experimental Results and Conclusions

The proposed algorithm has been tested on different test and real world signals. A significant and populated database of shapes (1024 images) and audio signals



**Fig. 4.** 1<sup>st</sup> row) Hammer shape; 2<sup>nd</sup> row) corresponding normalized differential chain code; 3<sup>rd</sup> row) transient component estimated using the proposed approach; 4<sup>th</sup> row) wavelet details at a fixed scale level of the transient component; 5<sup>th</sup> row) wavelet details at the same scale level of the differential normalized chain code

(300 signals) has been considered while for real shapes we have analysed cavities on the Roman Theatre of Aosta. Achieved results are quite encouraging for both shape analysis and audio processing, since the algorithm selects transient locations almost precisely. In all tests a spline 3/9 biorthogonal wavelet has been used and 5 dyadic scale levels have been considered. The parameter  $p$  in step 1 of the algorithm has been set equal to 0.1. The peculiarity of the algorithm is the fact that it tries to reconstruct the transient contribution even at scales where it is dominated by the tonal component (see rightmost figure, in Fig. 3). This allows us to avoid a pre-processing of the signal for detecting the tonal component. In this way it can be used in the processing of the tonal component since it allows to split the signal into distinct stationary pieces. The location of transients is not directly derived from the coarsest scale since they are subjected to transport. Then, it cannot correspond to the actual one in the time domain. For that reason, it is important to recover the contribution of transients at each scale. There are not objective measures for evaluating the final result. For that reason we will provide different examples that are able to show the potential of the method. In Fig. 3, a rectangle and star shapes have been considered. The detected corners are indicated and the wavelet details are compared. It is worth noticing that for differently oriented rectangles, different scale levels have been selected at step 5 of the algorithm, since the different tonal component. For the rectangle oriented at  $45^\circ$ , the selected scales have been  $s = 1.1, 3.1, 7.1, 11.2, 31.1$  while for the rectangle oriented at  $20^\circ$ , the scales are  $s = 1.1, 3.8, 5.7, 11.3, 31.1$ . Fig. 4 depicts two hammer shapes having a different orientation. They provide



**Fig. 5.** Castagnette (top), Glockenspiel (bottom left) and Xilofone (bottom right) test audio signals (top) and the corresponding three components.  $f_{tran}$  has been estimated using the proposed model, while the tonal component has been extracted by thresholding the DCT transform. The residual component is 1/100 of the original signal.

an oscillating normalized differential chain code but their significant corners have been recognized and classified as transient component. Some tests on audio and music signals have been also performed and some results are shown in Fig. 5. Also in this case there is a good detection. To give a measure of the result we used a DCT based techniques for detecting the tonal component and then we measure the energy of the residual. As it can be observed, it is low and it does not show significant contributions in correspondence to transient locations. It is a stochastic signal with a low power spectrum.

With regard to the reconstruction of the transient component, the faithfulness of the representation depends on the adopted basic atom. If for audio signals



we can have a very generic signal, for the normalized chain code, the basic signal is similar to a delta function. Hence, better results can be obtained if the algorithm is applied using the corresponding basic shape in the wavelet domain instead of the piecewise linear one — see right part of Fig. 3. Achieved results are encouraging and model's refinements will be investigated in the future.

## References

1. Bello, J.P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M., Sandler, M.B.: A Tutorial on Onset Detection in Music Signals. *IEEE Trans. on Speech and Audio Processing* 13(5), 1035–1047 (2005)
2. Bruni, V., Piccoli, B., Vitulano, D.: Wavelet time-scale Dependencies for Signal and Image Compression. In: Pan, Y., Chen, D.-x., Guo, M., Cao, J., Dongarra, J. (eds.) *ISPA 2005. LNCS*, vol. 3758, pp. 105–110. Springer, Heidelberg (2005)
3. Bruni, V., Vitulano, D.: Wavelet based Signal Denoising via Simple Singularities Approximation. *Signal Processing* 86, 859–876 (2006)
4. Daudet, L.: A Review on Techniques for the Extraction of Transients in Musical Signals. *Computer Music Modeling and Retrieval*, 219–232 (2005)
5. Daudet, L.: Sparse and Structured Decompositions of Signals with the Molecular Matching Pursuit. *IEEE Trans. on Audio, Speech and Language Processing* 14(5) (2006)
6. Daudet, L., Torresani, B.: Hybrid representations for audiophonic signal encoding. *Signal Processing, Special issue on Image and Video Coding Beyond Standards* 82(11), 1595–1617 (2002)
7. Duxbury, C., Davies, M.E., Sandler, M.B.: Separation Of Transient Information In Musical Audio Using Multiresolution Analysis Techniques. In: 4th Int. Workshop on Digital Audio Effects, DAFX 2001, Limerick (2001)
8. Duxbury, C., Chetry, N., Sandler, M., Davies, M.E.: An efficient two-stage implementation of Harmonic Matching Pursuit. In: *Proc. of EUSIPCO 2004(2004)*
9. Gonzales, R.C., Woods, R.E.: *Digital Image Processing*. Prentice-Hall, Englewood Cliffs (2002)
10. Grinboval, R., Bacry, E.: Harmonic Decomposition of Audio Signals with Matching Pursuit. *IEEE Trans. on Signal Processing* 51(1) (January 2005)
11. Mallat, S., Hwang, W.L.: Singularity Detection and Processing with Wavelet. *IEEE Trans. on Information Theory* 38, 617–643 (1992)
12. Mallat, S.: *A Wavelet Tour of Signal Processing*. Academic Press, London (1999)
13. Molla, S., Torresani, B.: Determining local transientness in audio signals. *IEEE Signal Processing Letters* 11(7), 625–628 (2004)
14. Tantibundhit, C., Boston, J.R., Li, C.C., Durrant, J.D., Shaiman, S., Kovacyk, K., El-Jaraoudi, A.: Speech Enhancement using Transient Speech Components. In: *Proc. of IEEE ICASSP 2006* (2006)