

# Determining local transientness of audio signals

Stéphane Molla and Bruno Torrèsani

## Abstract

We describe a new method for estimating the degree of “transientness” and “tonality” of a class of compound signals involving simultaneously transient and harmonic features. The key assumption is that both transient and tonal layers admit sparse expansions, respectively in wavelet and local cosine bases. The estimation is performed using particular form of entropy (or theoretical dimension) functions. We provide theoretical estimates on the behavior of the proposed estimators, as well as numerical simulations. Audio signal coding provides a natural field of application.

## Index Terms

audiophonic signal, transient, tonal, wavelet basis, local cosine basis, sparsity.

EDICS: 1.TFSR, 2.AUEA

## I. INTRODUCTION

Many generic signal classes feature significantly different “components”, such as transients, (locally) sinusoidal or harmonic “partials”, or stochastic-like components in sounds, or edges, textures, etc. in images. Detecting the presence of such components is one of the classical signal processing problems (see for example [1] and [2] and references therein for reviews.) Another interesting problem is to estimate whether a given portion of a signal is for example more transient than harmonic or periodic, or in other words to estimate “transientness” or “tonality” indices: quantitative measures of the local proportion of transient and tonal features in a signal. Such indices find immediate applications in several contexts, including the hybrid signal coders [3], [4], [2] which use different methods for encoding transient or tonal regions (and were the main motivation of this work), more general purpose hybrid models [5], or similar recent ideas in image coding [6], [7]. While there exist fairly standard tools for transient detection

Both authors are with LATP, CMI, 39 rue Joliot-Curie, 13453 Marseille cedex 13, France; email: molla@cmi.univ-mrs.fr; torresan@cmi.univ-mrs.fr. Work supported in part by the European Union’s Human Potential Programme, under contract HPRN-CT-2002-00285 (HASSIP)

or harmonic signal detection, the problem of quantitative measure of proportion does not seem to have received much attention.

We propose here simple criteria, based on transform coding ideas, for estimating such indices. The main idea is to use orthonormal bases in signal spaces which are significantly different from each other in the following sense: a given component has a sparse expansion in a given basis, while the others have dense expansions. Information theoretic criteria (we elaborate on the case of a variant of Shannon's entropy) therefore yield estimates for the indices.

We focus here on the case of transient and locally sinusoidal (or harmonic) layers in audio signals, using wavelet and local cosine bases. However, the approach we develop may be adapted to different signal layers (chirps for example), and to higher dimensions. We provide theoretical estimates for the behavior of transientness and tonality indices, and illustrate our results by numerical simulations and tests on real sounds.

## II. THEORETICAL ANALYSIS

We focus on the particular application to audio signals, and limit ourselves to transient and tonal features. Our starting point is the assumption that transient signals admit a sparse expansion in a wavelet basis (provided the wavelets have small enough time support), and that tonals admit a sparse expansion in local cosine basis (with smooth enough window function.) We are naturally led to consider a generic redundant "dictionary" made out of two such orthonormal bases, denoted by  $\psi_\lambda$  and  $w_\delta$  respectively (we refer to [8], [9] for detailed tutorials), and signal expansions of the form

$$x = \sum_{\lambda \in \Lambda} \alpha_\lambda \psi_\lambda + \sum_{\delta \in \Delta} \beta_\delta w_\delta + r = x_{tr} + x_{ton} + r, \quad (1)$$

where  $\Lambda$  and  $\Delta$  are (small, and this will be the main *sparsity* assumption) subsets of the index sets, termed *significance maps*. The nonzero coefficients  $\alpha_\lambda$  are assumed to be independent  $\mathcal{N}(0, \sigma_\lambda)$  random variables, and the nonzero coefficients  $\beta_\delta$  are assumed to be independent  $\mathcal{N}(0, \tilde{\sigma}_\delta)$  random variables.  $r$  is a residual signal, which is not sparse with respect to the two considered bases (we shall talk of *spread residual*), and is to be neglected or described differently.

Given a signal assumed for simplicity to be of the form (1), with unknown values of  $|\Delta|$  and  $|\Lambda|$  (the cardinalities of  $\Delta$  and  $\Lambda$  respectively), we are interested in finding estimates for the latter. More precisely,

we seek estimates for the “transientness” and “tonality” indices

$$I_{tr} = \frac{|\Lambda|}{|\Delta| + |\Lambda|}, \quad I_{ton} = \frac{|\Delta|}{|\Delta| + |\Lambda|}. \quad (2)$$

We propose a procedure close to the notions of *theoretical dimension* or  $\alpha$ -entropies, advocated by Wickerhauser (see [10] for a review.) Our approach is based upon the following heuristics. Consider a signal  $x$ , and expand it into an orthonormal basis. Estimating the “size” of  $x$  in this basis may clearly be done by counting the number of nonzero coefficients (the  $\ell^0$  norm of the sequence of coefficients), or the number of coefficients above some threshold. It has been shown [10] that alternative approaches are possible, including  $\ell^p$  norms (with  $p < 2$ ) or entropy, yielding comparable results. Considering a hybrid signal as in (1) neither its wavelet expansion nor its local cosine expansion will be sparse. However, since by assumption only very few coefficients  $\alpha_\lambda$  and  $\beta_\delta$  are nonzero, most wavelet coefficients  $\langle x, \psi_\lambda \rangle$  actually originate from the tonal part  $x_{ton}$ , and most local cosine coefficients  $\langle x, w_\delta \rangle$  originate from the transient part  $x_{tr}$ . Therefore, calculating these  $\ell^p$  norms or entropy from the wavelet coefficients  $\langle x, \psi_\lambda \rangle$  is expected to provide (approximately) estimates on the number of nonzero (or significant)  $\beta_\delta$  coefficients, and vice versa. We elaborate below on the specific case of the logarithmic dimensions, for which such a behavior may be proved. For the sake of simplicity, we shall work in this section in a finite dimensional context.

*Definition 1:* Given an orthonormal basis  $\mathcal{B} = \{e_n, n = 1, \dots, N\}$  of a given  $N$ -dimensional signal space  $\mathcal{E} \cong \mathbb{R}^N$ , define the logarithmic dimension of  $x \in \mathcal{E}$  in the basis  $\mathcal{B}$  by

$$\mathcal{D}_{\mathcal{B}}(x) = \frac{1}{N} \sum_{n=1}^N \log_2 (|\langle x, e_n \rangle|^2). \quad (3)$$

It follows from a simple calculation that in the framework of the signal models under consideration,

*Lemma 1:* Given an orthonormal basis  $\mathcal{B} = \{e_n, n = 1, \dots, N\}$ , assuming that the coefficients  $\langle x, e_n \rangle$  are  $\mathcal{N}(0, \sigma_n)$  random variables, one has

$$\mathbb{E} \{ \mathcal{D}_{\mathcal{B}}(x) \} = C + \frac{1}{N} \sum_{n=1}^N \log_2(\sigma_n^2), \quad (4)$$

where  $C = 1 + \gamma / \ln(2) \approx 1.832746177$  is a universal constant ( $\gamma \approx .5772156649$  being Euler’s constant.)

Returning to the model (1), and assuming that the coefficients  $\alpha_\lambda, \lambda \in \Lambda$  and  $\beta_\delta, \delta \in \Delta$  are respectively  $\mathcal{N}(0, \sigma_\lambda)$  and  $\mathcal{N}(0, \tilde{\sigma}_\delta)$  independent random variables, the coefficients  $a_\lambda = \langle x, \psi_\lambda \rangle$ ,  $b_\delta = \langle x, w_\delta \rangle$ , are zero-mean normal random variables, whose variance depends on whether  $\lambda \in \Lambda$  (or  $\delta \in \Delta$ ) or not. For

example,

$$\text{var}\{a_\lambda\} = \begin{cases} \sigma_\lambda^2 + \sum_{\delta \in \Delta} \tilde{\sigma}_\delta^2 |\langle \psi_\lambda, w_\delta \rangle|^2 & \text{if } \lambda \in \Lambda \\ \sum_{\delta \in \Delta} \tilde{\sigma}_\delta^2 |\langle \psi_\lambda, w_\delta \rangle|^2 & \text{if } \lambda \notin \Lambda \end{cases} \quad (5)$$

and we obtain, for the  $\Psi = \{\psi_\lambda\}$  basis

$$\begin{aligned} \mathbb{E}\{\mathcal{D}_\Psi(x)\} = & C + \frac{1}{N} \log_2 \left[ \prod_{\lambda \in \Lambda} \left( \sigma_\lambda^2 + \sum_{\delta \in \Delta} \tilde{\sigma}_\delta^2 |\langle \psi_\lambda, w_\delta \rangle|^2 \right) \right. \\ & \left. \times \prod_{\lambda' \notin \Lambda} \left( \sum_{\delta \in \Delta} \tilde{\sigma}_\delta^2 |\langle \psi_{\lambda'}, w_\delta \rangle|^2 \right) \right], \end{aligned} \quad (6)$$

and a similar expression for the logarithmic dimension  $\mathcal{D}_W(x)$  with respect to the  $W = \{w_\delta\}$  basis.

In the simpler case where  $\sigma_\lambda = \sigma, \forall \lambda \in \Lambda$  and  $\tilde{\sigma}_\delta = \tilde{\sigma}, \forall \delta \in \Delta$ , we introduce the Parseval weights

$$p_\lambda(\Delta) = \sum_{\delta \in \Delta} |\langle w_\delta, \psi_\lambda \rangle|^2, \quad \tilde{p}_\delta(\Lambda) = \sum_{\lambda \in \Lambda} |\langle w_\delta, \psi_\lambda \rangle|^2. \quad (7)$$

The following property is an immediate consequence of Parseval's formula (i.e. for all  $f, \sum_\lambda |\langle f, \psi_\lambda \rangle|^2 = \|f\|^2$ .)

*Lemma 2:* The Parseval weights satisfy

$$0 \leq p_\lambda(\Delta) \leq 1, \quad 0 \leq \tilde{p}_\delta(\Lambda) \leq 1.$$

Introducing the *relative redundancies* of the bases  $\Psi$  and  $W$  with respect to the significance maps

$$\epsilon(\Delta) = \sup_{\lambda \in \Lambda} p_\lambda(\Delta), \quad \tilde{\epsilon}(\Lambda) = \sup_{\delta \in \Delta} \tilde{p}_\delta(\Lambda), \quad (8)$$

we obtain simple estimates for the logarithmic dimension.

*Theorem 1:* With the above notations, assuming that the coefficients  $\{\alpha_\lambda, \lambda \in \Lambda\}$  and  $\{\beta_\delta, \delta \in \Delta\}$  are independent identically distributed  $\mathcal{N}(0, \sigma)$  and  $\mathcal{N}(0, \tilde{\sigma})$  normal variables respectively, and assuming  $r = 0$ , the following bounds hold

$$\begin{aligned} \mathbb{E}\{\mathcal{D}_\Psi(x)\} \geq & C + \frac{|\Lambda|}{N} \log_2(\sigma^2) \\ & + \log_2 \left( \prod_{\lambda' \notin \Lambda} (\tilde{\sigma}^2 p_{\lambda'}(\Delta))^{1/N} \right) \end{aligned} \quad (9)$$

$$\begin{aligned} \mathbb{E}\{\mathcal{D}_\Psi(x)\} \leq & C + \frac{|\Lambda|}{N} \log_2(\sigma^2 + \epsilon(\Delta)\tilde{\sigma}^2) \\ & + \log_2 \left( \prod_{\lambda' \notin \Lambda} (\tilde{\sigma}^2 p_{\lambda'}(\Delta))^{1/N} \right), \end{aligned} \quad (10)$$

with  $C = 1 + \gamma/\ln(2) \approx 1.832746177$ . Exchanging the roles of  $\Delta$  and  $\Lambda$ , a similar bound holds for the other logarithmic dimension  $\mathcal{D}_W(x)$ .

*Proof:* The proposition follows directly from the fact that in such a situation, equation (6) reduces to

$$\mathbb{E} \{ \mathcal{D}_\Psi(x) \} = C + \log_2 \left( \prod_{\lambda \in \Lambda} (\sigma^2 + \tilde{\sigma}^2 p_\lambda(\Delta))^{1/N} \times \prod_{\lambda' \notin \Lambda} (\tilde{\sigma}^2 p_{\lambda'}(\Delta))^{1/N} \right), \quad (11)$$

from Lemma 2 and the definition of  $\epsilon(\Delta)$ . ♠

This result may be understood and utilized as follows. First notice that the bounds in Equations (9) and (10) differ by  $|\Lambda| \log_2(1 + \epsilon(\Delta) \tilde{\sigma}^2 / \sigma^2) / N$ . Let us assume for a while that this term may be neglected (more on that below.) Then the behavior of  $\mathbb{E} \{ \mathcal{D}_\Psi(x) \}$  is essentially controlled by  $\log_2 \left( \prod_{\lambda' \notin \Lambda} (\tilde{\sigma}^2 p_{\lambda'}(\Delta))^{1/N} \right)$ . The behavior of this term is not easy to understand, but a first idea may be obtained by replacing  $p_{\lambda'}(\Delta)$  by its ‘ensemble average’  $\frac{1}{N} \sum_{\lambda=1}^N p_\lambda(\Delta) = \frac{1}{N} \sum_{\delta \in \Delta} \|w_\delta\|^2 = \frac{|\Delta|}{N}$ , which yields the approximate expression:

$$\mathbb{E} \{ \mathcal{D}_\Psi(x) \} \approx C + \frac{|\Lambda|}{N} \log_2(\sigma^2) + \left( 1 - \frac{|\Lambda|}{N} \right) \log_2 \left( \tilde{\sigma}^2 \frac{|\Delta|}{N} \right). \quad (12)$$

Therefore, if the ‘ $\Psi$ -component’ of the signal is sparse enough, i.e. if  $|\Lambda|/N$  is sufficiently small (compared with 1),  $\mathbb{E} \{ \mathcal{D}_\Psi(x) \}$  may be expected to behave as  $\log_2 \left( \tilde{\sigma}^2 \frac{|\Delta|}{N} \right)$ . Set

$$\hat{N}_\Psi(x) = 2^{\mathcal{D}_\Psi(x)}. \quad (13)$$

Replacing  $\mathcal{D}_\Psi(x)$  with its expectation, we see that  $\hat{N}_\Psi(x) \approx 2^C \tilde{\sigma}^2 \frac{|\Delta|}{N}$ , which yields an estimate (up to the multiplicative constant  $2^C \tilde{\sigma}^2 / N$ ) for the ‘size’ of the tonal component of the signal. Similarly, defining

$$\hat{N}_W(x) = 2^{\mathcal{D}_W(x)} \quad (14)$$

we obtain a similar estimate (up to the multiplicative constant  $2^C \sigma^2 / N$ ) for the ‘size’ of the  $\Psi$  component of the signal. Both  $\hat{N}_\Psi(x)$  and  $\hat{N}_W(x)$  are computable from the signals wavelet and local cosine expansions, and we finally consider their relative proportions, or ‘rates’

$$\hat{I}_{ton} = \frac{\hat{N}_\Psi}{\hat{N}_\Psi + \hat{N}_W}, \quad \hat{I}_{tr} = \frac{\hat{N}_W}{\hat{N}_\Psi + \hat{N}_W}, \quad (15)$$

which provide the desired estimates for the indices in Eq. (2).

A few comments are in order here.

- i.* The difference between the lower and upper bounds depends on the sparsity  $|\Lambda|/N$  of the  $\Psi$ -component and the relative redundancy parameters  $\epsilon(\Delta)$ . The latter actually describe the intrinsic

differences between the two considered bases. When the bases are significantly different, the relative redundancy may be expected to be small (notice that in any case, it is smaller than 1.)

- ii. The relative redundancy parameters  $\epsilon$  and  $\tilde{\epsilon}$  differ from the one which is generally considered in the literature, namely the *coherence*  $M[W \cup \Psi] = \sup_{b \in W, b' \in \Psi} |\langle b, b' \rangle|$  of the dictionary  $W \cup \Psi$  (see e.g. [11], [12], [13].) The latter is intrinsic to the dictionary, while the Parseval weights and corresponding  $\epsilon$  and  $\tilde{\epsilon}$  provide a finer information, as they also account for the signal models, via their dependence in the significance maps  $\Lambda$  and  $\Delta$ .
- iii. Precise estimates for  $\epsilon$  and  $\tilde{\epsilon}$  are difficult to obtain (numerical simulations yield values around 1/4.) More precise models for the significance maps  $\Delta$  and  $\Lambda$  could provide better understanding. In particular, structured models such as those described in [14] (implementing time persistence in  $\Delta$  and scale persistence in  $\Lambda$ ) are expected to yield smaller values for the relative redundancies than models featuring uniformly distributed significance maps.

Another interesting point is the sensitivity of such tools with respect to departures from the model, or noise. We show that results similar to the above ones still hold true in the presence of white noise, i.e. assuming that the residual  $r$  in (1) is a zero-mean Gaussian white noise. In such a situation, denoting by  $s^2$  the variance of the noise  $r$ , equation (6) becomes

$$\mathbb{E} \{ \mathcal{D}_\Psi(x) \} = C + \frac{1}{N} \log_2 \left[ \prod_{\lambda \in \Lambda} \left( \sigma_\lambda^2 + \sum_{\delta \in \Delta} \tilde{\sigma}_\delta^2 |\langle \psi_\lambda, w_\delta \rangle|^2 + s^2 \right) \right. \\ \left. \times \prod_{\lambda' \notin \Lambda} \left( \sum_{\delta \in \Delta} \tilde{\sigma}_\delta^2 |\langle \psi_{\lambda'}, w_\delta \rangle|^2 + s^2 \right) \right],$$

and a similar expression for the logarithmic dimension  $\mathcal{D}_W(x)$  with respect to the  $W = \{w_\delta\}$  basis. Hence, the approximate expression (12) becomes

$$\mathbb{E} \{ \mathcal{D}_\Psi(x) \} \approx C + \frac{|\Lambda|}{N} \log_2(\sigma^2 + s^2) \\ + \left( 1 - \frac{|\Lambda|}{N} \right) \log_2 \left( \tilde{\sigma}^2 \frac{|\Delta|}{N} + s^2 \right).$$

The discussion above (suitably adaptated) still holds as long as the signal energy  $\tilde{\sigma}^2 |\Delta|$  exceeds the noise energy  $s^2 N$ .

These estimates yield the following algorithm for estimating transientness and tonality indices for sparse hybrid signals  $x$  as given in (1).

- 1) Compute logarithmic dimensions  $\mathcal{D}_\Psi(x)$  and  $\mathcal{D}_W(x)$

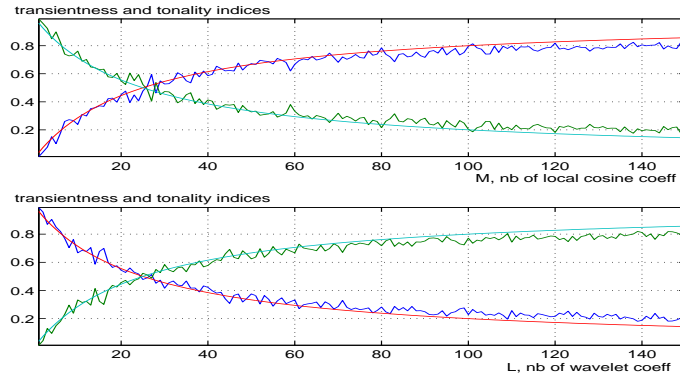


Fig. 1. Transientness and tonality estimates for the model (averaged over 10 realizations.) *Top*:  $L = 25$ , and  $M \in \{1, \dots, 150\}$ ; increasing curves:  $I_{ton}$  and  $\hat{I}_{ton}$ ; decreasing curves:  $I_{tr}$  and  $\hat{I}_{tr}$ . *Bottom*:  $M = 25$ , and  $L \in \{1, \dots, 150\}$ ; increasing curves:  $I_{tr}$  and  $\hat{I}_{tr}$ ; decreasing curves:  $I_{ton}$  and  $\hat{I}_{ton}$ .

- 2) Compute  $\hat{N}_{\Psi}(x)$  and  $\hat{N}_W(x)$  as in (13) and (14).
- 3) Compute the estimated rates as in (15).

### III. NUMERICAL RESULTS

We generated several realizations of the signal model (with  $r = 0$  first), with variable numbers  $L$  of wavelet atoms and fixed number  $M$  of local cosines and vice versa, and computed the estimated rates  $\hat{I}_{ton}$  and  $\hat{I}_{tr}$ , to be compared with the ground truth (2), i.e.  $I_{ton} = M/(M+L)$  and  $I_{tr} = L/(M+L) = 1 - I_{ton}$ . As may be seen from Figure 1 (which corresponds to averages over 10 realizations of the model), the estimated curves reproduce quite well the correct ones. Some discrepancies may be observed at the right hand side of the curves, where the sparsity assumptions are not valid any more, and the correction terms in (12) come into play. Observe that the curves cross precisely at the correct location  $M = L$ . The influence of the noise may be seen in Figure 2: a white noise, whose energy equals 30% of the signal's energy, has been added. The effect is what can be anticipated from (II), namely the presence of an additional noise term moves the experimental curves away from the theoretical ones.

Besides the numerical simulations above, the transientness and tonality indices have been tested on real audio signals, yielding very sensible results.<sup>1</sup> A first example, based upon a simple 'castanet' signal (6 s long, sampled at 44,100 kHz) is shown in Figure 3. A value for the transientness index and the tonality index was computed for all time frames (23 ms long.) Since  $I_{ton} = 1 - I_{tr}$ , only the transientness index is

<sup>1</sup>Additional material, including sound files, may be found at the web site <http://www.cmi.univ-mrs.fr/~torresan/papers/balance>.

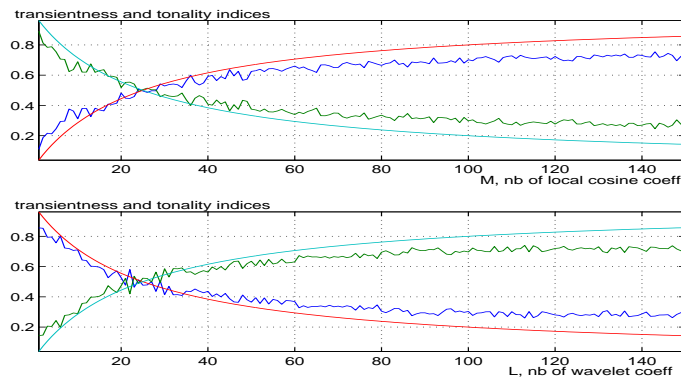


Fig. 2. Influence of white noise: Transientness and tonality estimates for the model (averaged over 10 realizations) with additional white noise. Same legends as before.

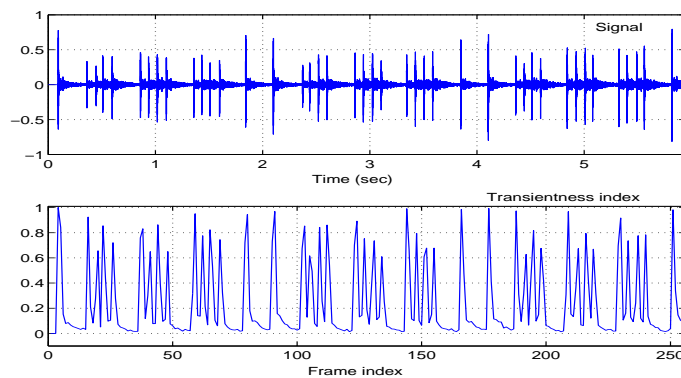


Fig. 3. Transientness index for the test 'castanet' signal. Signal (top) and transientness index (bottom.).

displayed for the sake of clarity. This signal is quite simple, as it essentially exhibits attacks followed by harmonic tones, and is thus a “perfect” test for the proposed approach. As may be seen from the bottom plot of Figure 3, all attacks are correctly captured, and the corresponding index is quite high. In between attacks, the transientness index is very low, which is also natural since the signal is essentially harmonic, thus sparsely represented by local cosine basis.

The second sound example displayed here is a more complex audio signal, extracted from a jazz recording (about 6 s. long, sampled at 44,100 kHz) which features “mixed” tonals and transients. The numerical results are displayed in Figure 4. Notice again that the “obvious” attacks of the signal have been captured by the method. A closer examination of the signal (using a “spectrogram type” representation, not shown here) shows that in the middle part of the signal (more precisely, between seconds 3 and 5), the harmonic content is stronger, which explains the lower average value of  $I_{tr}$  there. This illustrates the fact that  $I_{tr}$  really provides an estimate of the *proportion* of transients relative to tonals, rather than an absolute

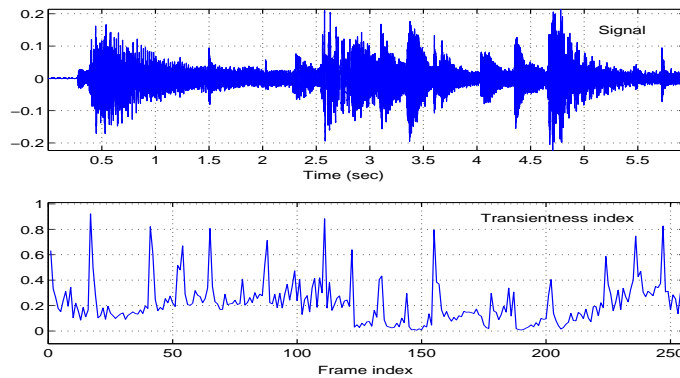


Fig. 4. Transientness index for the test ‘jazz’ signal. Signal (top) and transientness index (bottom.)

indicator of the presence of transient, such as the ones used in transient detection [4] for example.<sup>2</sup>

More numerical results, in the framework of the hybrid audio coding scheme developed in [3], will be given in a forthcoming publication [14].

#### IV. CONCLUSIONS

Sparsity of wavelet and local cosine signal representations may be exploited in order to estimate the relative amount of tonal and transient components present in the signal. This approach proves to be extremely effective in the context of hybrid audio signal coding [3], [15], and possesses a wider range of applications, including image coding [6].

The theoretical analysis we have presented is based on strong a priori assumptions on the signal (essentially, a hybrid model such as (1), with sparse significance maps  $\Lambda$  and  $\Delta$ .) While this sparsity assumption is completely necessary, the equality of variances may be relaxed; in that situation, the indices provide estimates on the proportion of energies  $\sigma^2|\Lambda|$  and  $\tilde{\sigma}^2|\Delta|$  of the two layers, rather than their size  $|\Lambda|$  and  $|\Delta|$ . This is the case for the numerical results on real signals, for which the variances of the two layers are not known.

Finally, let us simply mention that the approach may be extended to more than two layers, provided that the considered orthonormal bases are sufficiently different (in terms of their ‘Parseval weights’, see above) to allow the separation. Again, this may prove useful in the context of image coding, where new types of waveforms (e.g. curvelets) may be introduced.

<sup>2</sup><http://www.cmi.univ-mrs.fr/~torresan/papers/balance>.

## REFERENCES

- [1] X. Rodet and F. Jaillet, Detection and modeling of fast attack transients, in *Proc. ICMC*, La Habana, Cuba, 2001, pp. 30–33.
- [2] T. Verma, *A perceptually based audio signal model with application to scalable audio compression*, PhD Thesis, Stanford University (2000).
- [3] L. Daudet and B. Torr'esani, Hybrid representations for audiophonic signal encoding. *Signal Processing*, vol. 82(11), pp. 1595–1617, 2002.
- [4] S. Levine and J. O. Smith, A switched parametric and transform audio coder, in *Proc. ICASSP*, Phoenix, vol. 2, pp. 985–988, March 1999.
- [5] J. Berger, R. Coifman, and M. Goldberg, Removing noise from music using local trigonometric bases and wavelet packets. *J. Audio Eng. Soc.*, vol. 42(10), pp. 808–818, 1994.
- [6] F.G. Meyer, A.Z. Averbush and R.R. Coifman, Multilayered image representation: application to image compression. *IEEE Trans. on Image Processing*, vol. 11, pp. 1072–1080, 2002.
- [7] J. Romberg, M. Wakin and R. Baraniuk, Approximation and compression of piecewise smooth images using a wavelet/wedgelet geometric model. in *Proc. IEEE ICIP*, sept. 2003.
- [8] R. Carmona, W.L. Hwang, and B. Torr'esani, *Practical Time-Frequency Analysis: continuous wavelet and Gabor transforms, with an implementation in S*, Academic Press, San Diego, 1998.
- [9] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1998.
- [10] M. V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*. AK Peters, Boston, MA, USA, 1994.
- [11] D.L. Donoho and X. Huo, Uncertainty principles and ideal atomic decompositions. *IEEE Trans. Inf. Th.*, vol. 47(7), pp. 2845–2862, 2001.
- [12] M. Elad and A.M. Bruckstein, A generalized uncertainty principle and sparse representations. *IEEE Trans. Inf. Th.*, vol. 48(9), pp. 2558–2567, 2001.
- [13] R. Gribonval and M. Nielsen, Sparse representations in union of bases. Institut National de Recherches en Informatique et Automatique, IRISA Rennes, Tech. Rep. 1499, 2003.
- [14] S. Molla and B. Torr'esani, An Hybrid Audio scheme using hidden Markov models of waveforms. Submitted to *Appl. and Comp. Harm. Anal.* Preprint, Sept. 2003.
- [15] S. Molla and B. Torr'esani, Hidden Markov trees of wavelet coefficients for transient detection in audiophonic signals. To appear in *Annales de l'Université Blaise Pascal* A. Benassi Ed., 2003.