Harmonic Decomposition of Audio Signals With Matching Pursuit

Rémi Gribonval and Emmanuel Bacry

***Abstract*—We introduce a dictionary of elementary waveforms, called *harmonic atoms*, that extends the Gabor dictionary and fits well the natural harmonic structures of audio signals. By modifying the "standard" matching pursuit, we define a new pursuit along with a fast algorithm, namely, the fast harmonic matching pursuit, to approximate N-dimensional audio signals with a linear combination of M harmonic atoms. Our algorithm has a computational complexity of $\mathcal{O}(MKN)$, where K is the number of partials in a given harmonic atom. The decomposition method is demonstrated on musical recordings, and we describe a simple note detection algorithm that shows how one could use a harmonic matching pursuit to detect notes even in difficult situations, e.g., very different note durations, lots of reverberation, and overlapping notes.**

***Index Terms*—Audio signals, fundamental frequency extraction, Gabor atom, harmonic structure, matching pursuit, note detection, time–frequency analysis.**

I. Introduction

AUDIO signals contain superimposed structures such as transients and stationary parts. It has been noticed [22], [23], [26] that Gabor atoms

$$g_{s,\,u,\,\xi}(t) := \frac{1}{\sqrt{s}}\, w\left(\frac{t-u}{s}\right) e^{i2\pi\xi(t-u)} \tag{1}$$

provide a redundant family (*dictionary*) of elementary waveforms (*atoms*) that is well suited for decomposing such signals. However, given the strong harmonic content of most audio signals, it seems more natural to use a dictionary of *harmonic atoms*

$$h(t) := \sum_{k=1}^{K} c_k g_{s,\,u,\,\xi_k}(t), \qquad \|h\|^2 = \int |h(t)|^2\, dt = 1 \tag{2}$$

where $\xi_k \approx k\xi_0$, $1 \le k \le K$. Indeed, these elementary waveforms reflect well the prior knowledge about the structure of the signal. We define in this paper a modification of the matching pursuit algorithm [22] to decompose efficiently audio signals into linear combinations of such harmonic atoms.

Dictionaries of harmonic atoms are defined in Section II. In Section III, we recall the definition of the "standard" matching pursuit and introduce our modified matching pursuit, stating a convergence theorem (the proof is given in Appendix A). A fast implementation of this pursuit, namely, the fast harmonic matching pursuit, is described in more detail in Section V. Some applications of the harmonic matching pursuit are then described: A time–frequency representation is defined in Section VI, examples of the analysis of real audio signals are given in Section VII, and a simple note detection algorithm is experimented in Section VIII.

Manuscript received May 30, 2001; revised July 31, 2002. The associate editor coordinating the review of this paper and approving it for publication was Prof. Fredrik Gustafsson.

R. Gribonval is with IRISA-INRIA, French National Center for Computer Science and Control (INRIA), Rennes, France (e-mail: Remi.Gribonval@inria.fr).

E. Bacry is with the Center for Applied Mathematics (CMAP), École Polytechnique, Palaiseau, France (e-mail: emmanuel.bacry@polytechnique.fr).

Digital Object Identifier 10.1109/TSP.2002.806592

II. Dictionaries of Harmonic Atoms

A. Gabor and Harmonic Atoms

1) Gabor Atoms: Gabor atoms [see (1)] are obtained by dilating, translating, and modulating a mother window $w(t)$, which is generally real-valued, positive and of unit norm $\int |w(t)|^2\, dt = 1$. A Gabor atom $g_{s,\,u,\,\xi}(t)$ is located around time u with a duration of the order of s, and its Fourier transform $\widehat{g}_{s,\,u,\,\xi}(\omega)$ is centered at frequency ξ with a dispersion in frequency of the order of $1/s$.

2) Harmonic Atoms: Harmonic atoms [see (2)] are defined by their scale s, time u, frequency components $\xi_1 < \xi_2 < \cdots < \xi_K$, and by the complex coefficients $(c_k)_{k=1}^{K}$. A harmonic atom has the same localization in time as a Gabor atom, and its Fourier transform has essentially K peaks, located around frequencies ξ_k, $1 \le k \le K$, with a common width of the order of $1/s$.

Remark 1: Obviously, Gabor atoms are special cases of harmonic atoms, with $K = 1$. Moreover, real-valued Gabor atoms

$$g_{s,\,u,\,\xi,\,\phi}(t) := c_{s,\,\xi,\,\phi} w\left(\frac{t-u}{s}\right) \cos(2\pi\xi(t-u) + \phi) \tag{3}$$

are also harmonic atoms, with $K = 2$, $\xi_1 = -\xi$, and $\xi_2 = \xi$, and $c_{s,\,\xi,\,\phi}$ is a normalizing constant. In practical applications, we will also consider real-valued harmonic atoms, which are simply harmonic atoms with $\xi_{-k} = -\xi_k$, $1 \le k \le K$.

In the context of audio signals, it seems natural to "tune" the harmonic atoms in order to fit one of the main structure of these signals, namely, the *(almost) harmonicity* $\xi_k \approx k\xi_0$ between the frequency ξ_k of the kth *partial* $g_{s,\,u,\,\xi_k}$ and the *fundamental frequency* ξ_0 [7], [8]. Taking into account the spectral width (of the order of $1/s$) of the kth partial $g_{s,\,u,\,\xi_k}$, the (almost)harmonicity can be written[1] $|\xi_k - k\xi_0| \le A/s$, $1 \le k \le K$, with $A \approx 1$.

[1]If some prior knowledge is available, one could improve the analysis presented here by using a more precise model of the (almost) harmonicity of the partials. For example, if the signal is a recording of a piano piece, the relation $\xi_k \approx k\xi_0\sqrt{1 + bk^2}$ [10] makes it possible to define more adapted harmonic atoms.

B. Harmonic Subspace—Quasi-Orthogonality of the Partials

The set of all harmonic atoms at scale s, time u, and with frequency components $\xi_1 < \xi_2 < \cdots < \xi_K$ is exactly the unit sphere $\mathcal{S}_{s,\,u,\,\xi_1,\,...,\,\xi_K}$ in the subspace of $\boldsymbol{L}^2(\mathbb{R})$

$$\mathcal{V}_{s,\,u,\,\xi_1,\,...,\,\xi_K} := \text{span}\,\{g_{s,\,u,\,\xi_k},\, 1 \le k \le K\} \tag{4}$$

which will be referred to as a *harmonic subspace*. When it is possible, it is helpful to specify a range of possible fundamental frequencies

$$\xi_0^{\min}(s,\,u) \le \xi_0 \le \xi_0^{\max}(s,\,u). \tag{5}$$

This may come from *a priori* knowledge on the audio signal. For some technical reason that will become clearer later on, we need the partials $\{g_{s,\,u,\,\xi_k},\, 1 \le k \le K\}$ to be *quasiorthogonal*, that is to say for some $0 < \delta \le 1$

$$|\langle g_{s,\,u,\,\xi_k},\, g_{s,\,u,\,\xi_l}\rangle| = \left|\widehat{w^2}(s(\xi_l - \xi_k))\right| \le \frac{1-\delta}{\sqrt{K^2 - K}},\quad k \ne l. \tag{6}$$

It is not difficult to check that this is satisfied if $\xi_0^{\min}(s,\,u) \ge B/s$ for some constant $B = B(\delta,\,K,\,w)$.

C. Gabor and Harmonic Dictionaries

1) Gabor Dictionaries: The *Gabor dictionary* is the set $\mathcal{D}_g = \{g_{s,\,u,\,\xi},\, (s,\,u,\,\xi) \in \Gamma_g = \mathbb{R}_+ \times \mathbb{R}^2\}$ of Gabor atoms at every scale $s > 0$, time location $u \in \mathbb{R}$ and frequency $\xi \in \mathbb{R}$.

2) Harmonic Dictionaries: A *harmonic dictionary* $\mathcal{D}_h$ is an extension of the Gabor dictionary $\mathcal{D}_g$

$$\mathcal{D}_h := \mathcal{D}_g \cup \bigcup_{(s,\,u,\,\xi_1,\,...,\,\xi_K) \in \Gamma_h} \mathcal{S}_{s,\,u,\,\xi_1,\,...,\,\xi_K}$$

for some set of indices Γ_h. We will use the notation $\Gamma = \Gamma_g \cup \Gamma_h$, and $\gamma \in \Gamma$ will denote the index of either a Gabor atom $(s,\,u,\,\xi)$ or a harmonic subspace $(s,\,u,\,\xi_1,\,\ldots,\,\xi_K)$. Notice that due to the constraint (5), not all Gabor atoms $g \in \mathcal{D}_g$ lie in some $\mathcal{S}_\gamma$, $\gamma \in \Gamma_h$.

III. Standard Matching Pursuit

The matching pursuit [22] is a greedy algorithm very similar to the projection pursuit introduced in statistics [11], [19]. Given a complete dictionary $\mathcal{D}$, i.e., a redundant family of unit vectors in a Hilbert space $\mathcal{H}$ such that $\overline{\text{span}}(\mathcal{D}) = \mathcal{H}$, and an arbitrary number M, it decomposes a signal $s(t)$ into a residual term $R_M(t)$ and a linear combination of M atoms chosen among $\mathcal{D}$

$$s(t) = \sum_{m=1}^{M} \lambda_m g_m(t) + R_M(t)$$

with the essential *energy conservation* property

$$\|s\|^2 = \sum_{m=1}^{M} \lambda_m^2 + \|R_M\|^2.$$

The strong convergence $\lim_{M\to\infty} \|R_M\| = 0$ was proved by Jones [21] and shows that one can get as good an approximation to $s(t)$ as wanted.

A. Standard Matching Pursuit

From a decomposition of the signal into $M - 1 \ge 0$ atoms

$$s(t) = \sum_{m=1}^{M-1} \langle R_{m-1},\, g_m\rangle\, g_m(t) + R_{M-1}(t)$$

one gets an M-atom decomposition in the following way.

1) Compute $|\langle R_{M-1},\, g\rangle|$ for all $g \in \mathcal{D}$.
2) Select a (near) best atom of the dictionary

$$|\langle R_{M-1},\, g_M\rangle| \ge \rho \sup_{g\in\mathcal{D}} |\langle R_{M-1},\, g\rangle|$$

where $0 < \rho \le 1$ is some number independent of M.
3) Compute the new residual

$$R_M(t) := R_{M-1}(t) - \langle R_{M-1},\, g_M\rangle g_M(t). \tag{7}$$

B. Standard Matching Pursuit in Harmonic Dictionaries

With harmonic dictionaries, one can write

$$\begin{aligned}\sup_{h\in\mathcal{D}_h} |\langle R_{M-1},\, h\rangle| &= \sup_{\gamma\in\Gamma} \sup_{h\in\mathcal{S}_\gamma} |\langle R_{M-1},\, h\rangle| \\ &= \sup_{\gamma\in\Gamma} \left\|P_{\mathcal{V}_\gamma} R_{M-1}\right\|\end{aligned}$$

where $P_{\mathcal{V}}$ is the orthonormal projection onto $\mathcal{V}$. Consequently, the standard matching pursuit can take the following form.

1) Compute $\|P_{\mathcal{V}_\gamma} R_{M-1}\|$ for all $\gamma \in \Gamma$.
2) Select a (near) best harmonic subspace

$$\|P_{\mathcal{V}_M} R_{M-1}\| \ge \rho \sup_{\gamma\in\Gamma} \left\|P_{\mathcal{V}_\gamma} R_{M-1}\right\|.$$

3) Compute the new residual as in (7) with

$$\begin{aligned} h_M(t) &:= \frac{P_{\mathcal{V}_M} R_{M-1}(t)}{\|P_{\mathcal{V}_M} R_{M-1}\|} \\ \langle R_{M-1},\, h_M\rangle &:= \|P_{\mathcal{V}_M} R_{M-1}\|. \end{aligned}$$

This formulation shows that no exhaustive search over the parameters $(c_k)_{k=1}^K$ is needed for the optimization of a harmonic atom. In particular, following Remark 1, when the dictionary consists of real-valued Gabor atoms [see (3)], the phase ϕ_m can be automatically optimized [22], [4], [14], [16].

IV. Approximate and Weak Harmonic Matching Pursuit

At each step of the standard matching pursuit described above, one needs to compute $\|P_{\mathcal{V}_\gamma} R_{M-1}\|$ for every subspace $\mathcal{V}_\gamma$, $\gamma \in \Gamma$, as well as the exact projection $P_{\mathcal{V}_M} R_{M-1}(t)$ for the selected subspace. For Gabor atoms ($K = 1$), this is easily done as $P_{\mathcal{V}_\gamma} R_{M-1}(t) = \langle R_{M-1},\, g_\gamma\rangle g_\gamma(t)$. A fast and exact computation is also possible with real-valued Gabor atoms ($K = 2$) [4], [14], [16].

For general harmonic dictionaries, computing $\|P_{\mathcal{V}_\gamma} R_{M-1}\|$ for every subspace $\mathcal{V}_\gamma$, $\gamma \in \Gamma$ is time consuming and, from a practical point of view, makes the standard matching pursuit unusable. In the next section, we describe how the quasiorthogonality of the partials [see (6)] along with some recent results on the convergence of *approximate weak greedy algorithms* [18]

can be used to define a modified pursuit that avoids these computations.

A. Modified Harmonic Matching Pursuit

Thanks to the quasiorthogonality condition [see (6)], the set of Gabor atoms $\{g_k = g_{s,u,\xi_k},\ 1 \le k \le K\}$ used to define $\mathcal{V}_\gamma$ in (4) is *nearly* an orthonormal basis of $\mathcal{V}_\gamma$. The *modified matching pursuit* is a standard matching pursuit acting *as if* the partials were *exactly* an orthonormal basis of the corresponding harmonic subspace.

If the partials were orthogonal, one would have

$$\left\|P_{\mathcal{V}_\gamma} R_{M-1}\right\|^2 = \sum_{k=1}^{K} \left|\langle R_{M-1},\, g_k\rangle\right|^2$$

and

$$P_{\mathcal{V}_\gamma} R_{M-1}(t) = \sum_{k=1}^{K} \langle R_{M-1},\, g_k\rangle\, g_k(t).$$

The modified matching pursuit is thus defined as follows.

1) Compute for all γ the *correlation function*

$$Q_{\mathcal{V}_\gamma}(R_{M-1}) := \sum_{k=1}^{K} \left|\langle R_{M-1},\, g_k\rangle\right|^2. \tag{8}$$

2) Select a harmonic subspace $\mathcal{V}_M$ such that

$$Q_{\mathcal{V}_M}(R_{M-1}) \ge \rho_M^2 \sup_{\gamma\in\Gamma} Q_{\mathcal{V}_\gamma}(R_{M-1}) \tag{9}$$

where $0 \le \rho_M \le 1$ may depend on M; see Theorem 1.

3) Compute the new residual according to

$$R_M(t) := R_{M-1}(t) - \lambda_M h_M(t) \tag{10}$$

where

$$\lambda_M h_M(t) := \sum_{k=1}^{K} \langle R_{M-1},\, g_{M,k}\rangle\, g_{M,k}(t) \tag{11}$$

with the notation $g_{m,k}(t) = g_{s_m,u_m,\xi_{m,k}}(t)$.

After M steps, this modified matching pursuit provides a decomposition of an audio signal as

$$\begin{aligned} s(t) &= \sum_{m=1}^{M} \lambda_m h_m(t) + R_M(t) \\ &= \sum_{m=1}^{M}\sum_{k=1}^{K} c_{m,k} g_{m,k}(t) + R_M(t). \end{aligned} \tag{12}$$

B. Convergence—Approximate and Weak Pursuit

The conditions for the convergence of the so-defined pursuit are given in the following theorem (the proof is in Appendix A).

Theorem 1: Assume that the harmonic dictionary satisfies the quasiorthogonality condition [see (6)]. Let $\{\rho_m\}_{m\ge1}$ with $0 \le \rho_m \le 1$ such that

$$\sum \frac{\rho_m}{m} = +\infty. \tag{13}$$

Then, the residual obtained with the modified matching pursuit [i.e., see (8)–(11)] converges strongly to zero.

Thus, the so-obtained matching pursuit is *weak* in the sense that the choice (9) of a "good" harmonic atom can be much *weaker* than in a standard matching pursuit. Indeed, one is not restricted to $\rho_m \ge \rho > 0$, and we will see in Section V an example where $\rho_m \in \{0, 1\}$, in which case, convergence is guaranteed if $\sum_m \rho_m/m = \sum_{m|\rho_m=1} 1/m = \infty$.

Moreover, it is *approximate* in the sense that, at each step, $\lambda_m h_m(t)$ defined by (11) is only an *approximation* of $\langle R_{m-1},\, h_m\rangle h_m(t)$, and the energy conservation is approximate as well, i.e.,

$$\|s\|^2 \simeq \sum_{m=1}^{M} |\lambda_m|^2 + \|R_M\|^2 \simeq \sum_{m=1}^{M}\sum_{k=1}^{K} |c_{m,k}|^2 + \|R_M\|^2 \tag{14}$$

as long as δ [defined by (6)] satisfies $\delta \simeq 1$. The proof and the precise formulation of this last equation can be found in Appendix B.

Let us now describe a fast implementation of this modified harmonic matching pursuit, namely, the fast harmonic matching pursuit.

V. Fast Harmonic Matching Pursuit

A. Main Principles

The main idea of the fast harmonic matching pursuit is that it is possible to select the "best" harmonic subspace $\mathcal{V}_m$ in a finite *sub-dictionary* $\mathcal{D}_m \subset \mathcal{D}_h$

$$Q_{\mathcal{V}_m}(R_{m-1}) := \max_{\gamma|\mathcal{S}_\gamma\subset\mathcal{D}_m} Q_{\mathcal{V}_\gamma}(R_{m-1}). \tag{15}$$

By choosing the sub-dictionaries $\{\mathcal{D}_m\}_{m\ge1}$ *much smaller* than the whole harmonic dictionary $\mathcal{D}_h$, we decrease the numerical complexity. By using (13), we can indeed construct small sub-dictionaries without loosing the convergence of the pursuit. The general principle is the following.

1) Initialization: At some steps $m_1 = 1 < m_2 < \cdots < m_p < \cdots$, the finite sub-dictionary $\mathcal{D}_{m_p}$ is *initialized* so that it satisfies $\max_{\gamma|\mathcal{S}_\gamma\subset\mathcal{D}_{m_p}} Q_{\mathcal{V}_\gamma}(R_{m-1}) = \max_{\gamma|\mathcal{S}_\gamma\subset\mathcal{D}_h} Q_{\mathcal{V}_\gamma}(R_{m-1})$.

2) Update: At the intermediate steps $m \in]m_p,\, m_{p+1}[$, the sub-dictionary $\mathcal{D}_m \subset \mathcal{D}_{m-1}$ is *updated* by removing some harmonic subspaces from $\mathcal{D}_{m-1}$.

B. Convergence

From the brief description above, one can easily show that at each step, the selected harmonic subspace $\mathcal{V}_m$ satisfies the "pessimistic" estimate $Q_{\mathcal{V}_m}(R_{m-1}) \ge \rho_m \sup_{\mathcal{S}_\gamma\subset\mathcal{D}_h} Q_{\mathcal{V}_\gamma}(R_{m-1})$ with

$$\rho_m := \begin{cases} 1, & m \in \{m_p,\, p \ge 1\} \\ 0, & m \notin \{m_p,\, p \ge 1\}. \end{cases}$$

Hence, if $\sum_m \rho_m/m = \sum_p (m_p)^{-1} = +\infty$, then $\{\rho_m\}_{m\ge1}$ is an acceptable "weakness" sequence (according to Theorem 1), and the pursuit will be convergent.

C. Adaptive Sub-Dictionaries of Local Maxima

Let us describe how $\mathcal{D}_{m_p}$ is initialized and how $\mathcal{D}_m$ is updated from $\mathcal{D}_{m-1}$.

1) Initialization: At each step m_p, we detect the *local maxima* of the function

$$u \mapsto \sum_{k=1}^{K} \sup_{|\xi_k - k\xi_0| \leq A/s,} \left|\langle R_{m_p-1},\, g_{s,u,\xi_k}\rangle\right|^2 \qquad (16)$$

for every value of s and ξ_0, as well as the local maxima of the function

$$\xi_0 \mapsto \sum_{k=1}^{K} \sup_{|\xi_k - k\xi_0| \leq A/s,} \left|\langle R_{m_p-1},\, g_{s,u,\xi_k}\rangle\right|^2 \qquad (17)$$

for every value of s and u and keep the location of the N_p largest (the choice of N_p will be discussed in a moment). This corresponds to keeping only the local maxima[2] for which the correlation is above some *threshold* η_p.

2) Update: The same threshold is used to update $\mathcal{D}_m$ for $m \in]m_p,\, m_{p+1}[$: Once $h_m(t) \in \mathcal{D}_m$ has been selected, one recomputes the correlation of the new residual with the subspaces of $\mathcal{D}_m$, and the threshold η_p is applied to obtain $\mathcal{D}_{m+1}$.

The next initialization step m_{p+1} occurs when the process of throwing away subspaces from $\mathcal{D}_{m_p}$ has emptied it.

D. Fast Matching Pursuit Algorithm

When one deals with a finite but high-dimensional signal of N samples, the standard discretization $\mathcal{D}_g^d$ of the Gabor dictionary contains $\mathcal{O}(N \log N)$ Gabor atoms.

Let us describe in details the implementation and the numerical complexity of the fast harmonic matching pursuit.

1) Initialization Steps $m \in \{m_p,\, p \geq 1\}$*:*

1) [$\mathcal{O}(N \log^2 N)$] Compute $\langle R_{m-1},\, g\rangle$ for every Gabor atom $g \in \mathcal{D}_g^d$. This is equivalent to computing several short time Fourier transforms (STFTs) based on windows at each possible scale, which is done using a fast algorithm (FFT or direct convolution).
2) [$\mathcal{O}(KN \log N)$] Compute, for every discrete $(s,\, u)$ and $\xi_0 \in [\xi_0^{\min}(s,\, u),\, \xi_0^{\max}(s,\, u)]$

$$\sum_{k=1}^{K} \sup_{|\xi_k - k\xi_0| \leq A/s,} \left|\langle R_{m-1},\, g_{s,u,\xi_k}\rangle\right|^2. \qquad (18)$$

There is at most $\mathcal{O}(N \log N)$ such discrete values of $(s,\, u,\, \xi_0)$.
3) [$\mathcal{O}(N \log N)$] Detect the local maxima.
4) [$\mathcal{O}(N \log^2 N)$] Sort the local maxima and threshold.

2) Updates for $m_p \leq m < m_{p+1}$*:*

1) [$\mathcal{O}(N_p)$] Select $(s_m,\, u_m,\, \xi_{m,0})$ of the "best" subspace, and set $\xi_{m,k} := k\xi_{m,0}$, $1 \leq k \leq K$.
2) [$\mathcal{O}(K)$] (Optional) Perform a Newton interpolation to get a fine estimate for $1 \leq k \leq K$ of

$$\xi_{m,k} := \arg \max_{\xi \in [k\xi_{m,0} - A/s,\, k\xi_{m,0} + A/s]} \left|\langle R_{m-1},\, g_{s_m,u_m,\xi}\rangle\right|.$$

3) [$\mathcal{O}(KN)$] Update the residual according to (10).
4) [$\mathcal{O}(K^2 N_p)$] Update the inner products of the useful Gabor atoms

$$\langle R_m,\, g\rangle = \langle R_{m-1},\, g\rangle - \sum_{k=1}^{K} \langle R_{m-1},\, g_{m,k}\rangle \langle g_{m,k},\, g\rangle. \qquad (19)$$

There are at most KN_p useful Gabor atoms: K in each subspace of $\mathcal{D}_m$. Each inner product $\langle g_{m,k},\, g\rangle$ can be computed in $\mathcal{O}(1)$ with an analytic formula (see e.g., [22]).
5) [$\mathcal{O}(KN_p)$] Recompute the correlations [see (18)] for every discrete $(s,\, u,\, \xi_0)$ associated with some subspace in $\mathcal{D}_m$.
6) [$\mathcal{O}(N_p)$] Eliminate from $\mathcal{D}_{m+1}$ the subspaces whose correlation has fallen below η_p.

E. Computational Complexity and Convergence

In practice we choose a *constant size* of $\mathcal{D}_{m_p}$

$$(K + \log N) \log N \ll N_p := N_0 \ll N/K.$$

The number of steps necessary to empty each sub-dictionary $\mathcal{D}_{m_p}$ is at least its size, i.e., $m_{p+1} - m_p \leq N_0$. Because local maxima of the correlation function have a tendency to be almost orthogonal one to another, only few subspaces are removed from $\mathcal{D}_m$ at each step, hence it is reasonable to assume that $m_{p+1} - m_p \geq \alpha N_0$ for some $\alpha > 0$. As a result, the computational complexity of M iterations of this pursuit is at most $\mathcal{O}((M/N_0)N(K + \log N) \log N) + \mathcal{O}(MKN(1 + N_0/N))$ that is to say

$$\mathcal{O}(MKN).$$

The convergence $\|R_m\| \to 0$ follows from Theorem 1 and the fact that $m_p \leq N_0 \times p \Rightarrow \sum_p (m_p)^{-1} = +\infty$.

VI. Time—Frequency Representation

The harmonic matching pursuit described in the previous sections allows one to decompose a signal $s(t)$ as the sum of a residual term and of a linear combination of an arbitrary number M of harmonic atoms, i.e.,

$$s(t) = \sum_{m=1}^{M} \lambda_m h_m(t) + R_M(t).$$

One fundamental property of this decomposition is that it satisfies the approximate energy conservation law

$$\|s\|^2 \simeq \sum_{m=1}^{M} |\lambda_m|^2 + \|R_M\|^2$$

[2]It has been observed [27] that local maxima of correlation functions such as (16) and (17) are likely to correspond to signal features. This is a desirable fact because it shows that for $m \in]m_p,\, m_{p+1}[$, the harmonic atom $h_m \in \mathcal{D}_m \subset \mathcal{D}_{m_p}$ will likely be a feature *of the signal* rather than an artefact of the matching pursuit, as was sometimes the case with the standard matching pursuit [6], [17], [20].

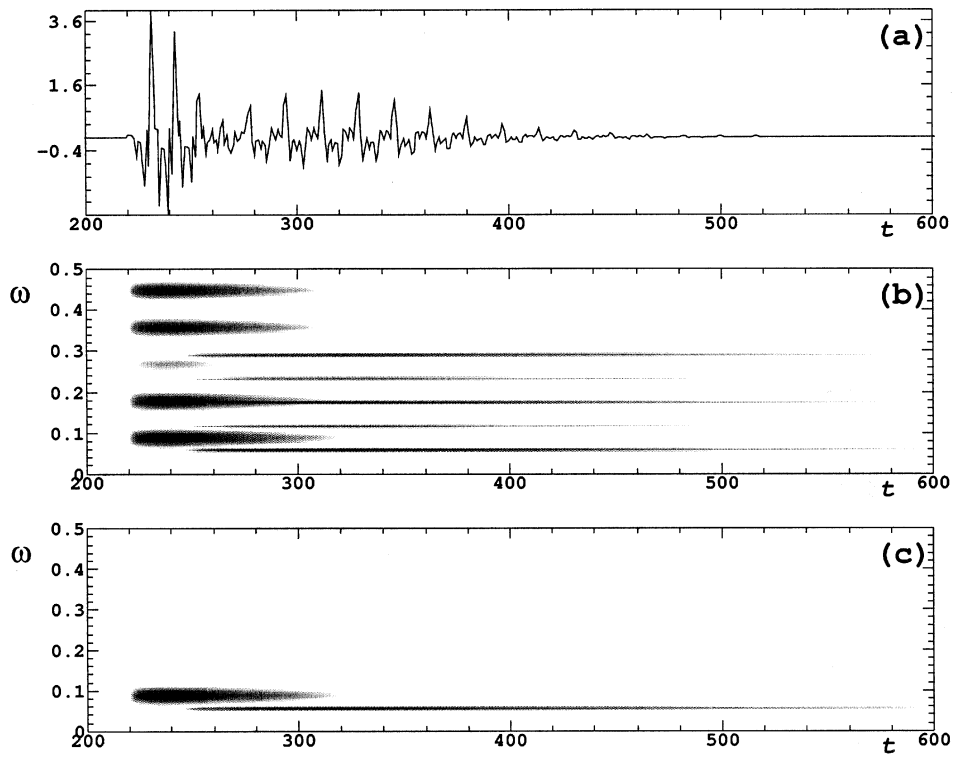

Fig. 1. Time–frequency representation using the harmonic matching pursuit. The analyzed signal is a simple sum of two harmonic atoms (with $K = 5$) using the asymmetric window $w(t)$ defined by (23) and displayed in Fig. 2. The first atom corresponds to $s_1 = 128$, $u_1 = 220$, and $\xi_{1,k} = 0.09k$ and the second one to $s_2 = 512$, $u_2 = 244$, and $\xi_{2,k} = 0.058k$. A $M = 2$ step harmonic matching pursuit is performed. (a) Graph of the analyzed signal. (b) Energy density $E_2[s](t, \omega)$ [see (20)]. (c) Reduced energy density $E_2^{(r)}[s](t, \omega)$ [see (21)].

(let us recall that the λ_ms are complex numbers and that $||h_m|| = 1$). Moreover, each harmonic component corresponds to a linear combination of Gabor atoms

$$\lambda_m h_m(t) = \sum_{k=1}^{K} c_{m,k} g_{m,k}(t),$$

and consequently

$$||s||^2 \simeq \sum_{m=1}^{M} \sum_{k=1}^{K} |c_{m,k}|^2 + ||R_M||^2.$$

Let us note that from a numerical point of view, these energy conservation relations [see (14) and its precise formulation in Appendix B] can be considered, in a very good approximation, to be exact equalities.

Following the usual representation used for the standard matching pursuit [22], we choose to represent each Gabor atom $g_{s,u,\xi}(t)$ in the (t, ω) time–frequency half-plane by its Wigner—Ville distribution [9] $WV[g_{s,u,\xi}](t, \omega)$. The energy density $E_M[s](t, \omega)$ of the signal $s(t)$ in the time-frequency half-plane at the step M of the pursuit is then naturally defined by

$$E_M[s](t, \omega) = \sum_{m=1}^{M} \sum_{k=1}^{K} |c_{m,k}|^2 WV[g_{m,k}](t, \omega). \qquad (20)$$

Fig. 1 illustrates the so-obtained time-frequency representation using the asymmetric window $w(t)$ defined by equation (23) and displayed in Fig. 2. It shows the energy density E_M corresponding to a signal made of two harmonic atoms (with $K = 5$) starting, respectively, at time $u_1 = 220$ and $u_2 = 244$. The black spots in Fig. 1(b) represent the Wigner–Ville distributions of the corresponding Gabor atoms. The two lower spots correspond to the fundamental frequencies of the two harmonic atoms and the other ones to their harmonics. In this case, since there are only two harmonic atoms, it is of course very easy to relate which spots belong to which harmonic atom. However, for real audio signals, this can get rather complicated and make this representation very hard to "read." For the sake of simplicity, we will use a "reduced" version of this representation, consisting of a representation of only the first partial of each harmonic atom. This reduced energy density $E_M^{(r)}[s](t, \omega)$ is then simply defined as

$$E_M^{(r)}[s](t, \omega) = \sum_{m=1}^{M} |\lambda_m|^2 WV[g_{m,1}](t, \omega) \qquad (21)$$

and it is illustrated in Fig. 1(c).

In the next section, we illustrate the harmonic matching pursuit on a real audio signal.

VII. Harmonic Matching Pursuit of a Real Audio Signal

Sound signals are asymmetric in time. They often consist in a short transient part (e.g., the attack of the sound) followed by a

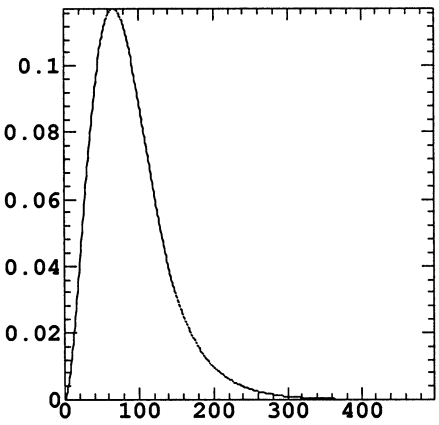

Fig. 2. FoF window. The graph of the asymmetric window $w(t)$ [see (23)] with $\beta = 4\pi$ and $\alpha = 10^5$.

stationary part which eventually slowly fades out (e.g., the sustained/decay parts of the sound). Consequently, as shown in a previous work [17], if a symmetrical window $w(t)$ is used, the matching pursuit will often pick up, as the most energetic atom, an atom that overlaps the actual starting time location of the transient. It results in "creation of energy" just before this transient. Thus, for instance, if the signal consists in a succession of notes played by an instrument, the matching pursuit time–frequency representation will display the energy of each note as if it was starting before its actual starting time location. This results in a pre-echo effect in the resynthesized audio signal:

$$s_M(t) = \sum_{m=1}^{M} \lambda_m h_m(t). \tag{22}$$

As suggested in [17] and [20], in order to avoid creation of energy, one could use a high-resolution matching pursuit. However, it slows down the pursuit quite a bit. In the case of audio signals, since the time asymmetry is basically always the *same* (e.g., the transients generally come before the stationary parts), the pre-echo effect can be taken care of (as shown in Fig. 3) by simply using an asymmetric window that reproduces a generic transient followed by a generic slowly decaying part [14], [15]. In order to keep a fast algorithm, one has to choose a window which enables an analytic formula of the inner product of two atoms [see (19)] [22]. For that purpose, we chose the FoF function [24], which is defined by

$$w(t) = \begin{cases} 0.5 * C(1 - \cos(\beta t))e^{-\alpha t}, & \text{for } 0 \le t < \dfrac{\pi}{\beta} \\ Ce^{-\alpha t}, & \text{for } \dfrac{\pi}{\beta} \le t < 1 \\ 0, & \text{otherwise} \end{cases} \tag{23}$$

where C is a normalization factor, β allows one to adjust the size of the transient, and α is the damping factor (let us note that α will be chosen so that the discontinuity at $t = 1$ is of the order of the numerical noise). The analytical formulas of the inner product of two such atoms are rather complicated and can be found in [1] and [3]. In the following, we will always choose $\beta = 4\pi$ and $\alpha = 10^5$. The graph of this window is displayed on Fig. 2. The computation of $\langle R_{m-1}, g \rangle$ at the initialization step are performed based on the FFT; however, as suggested in [15], they may be sped up by using recursive filters.

The harmonic matching pursuit using this asymmetric window is illustrated on Fig. 3. It has been performed on a real audio signal which consists in an 11-note melodic recording of a clarinet [5]. Fig. 3(c) shows the residual R_M for $M = 100$. It shows that 100 harmonic atoms are enough to capture most of the energy of the original signal (the relative L^2 error is of the order of 9 dB). Moreover, as it can be seen in Fig. 3(b), each note (with all its harmonic structure) is captured by very few atoms [to make this figure easier to read, each note has been indexed by order of appearance from the first one (1) to the last one (11)]. Moreover, the use of the asymmetric window (23) prevented any pre-echo, i.e., the beginning of the notes can be detected very precisely looking at the Fig. 3(b).

Thus, it seems very natural to use this harmonic decomposition to build a note detection algorithm. In the following section, we elaborate what could be the basis of such an algorithm.

VIII. Note Detection Algorithm Using Harmonic Matching Pursuit Decomposition

The note detection algorithm we describe here is basic and it should not be used as it is for note detection applications. The purpose of this section is to show that, although it is a first naive version of what should be a more elaborate algorithm, it can detect notes successfully even in some very difficult situations, e.g., very different note durations, lots of reverberation, etc. Moreover, let us point out that, apart from the frequency range [see (5)], the only prior information on the audio signal that is implicitly used by the algorithm is that the pitches of the notes do not change significantly through time. There is no prior information on what instruments are playing, how they are tuned, how many notes can be played at the same time, what type of music is played, etc.

A. Note Detection Algorithm

The basic idea of the algorithm is that the most energetic harmonic atoms are good candidates for notes. Given such an atom, the algorithm first evaluates what the fundamental frequency ω of the corresponding note is, based on the simplifying assumption that it corresponds to the most energetic partial. Then, it computes (using all the atoms of the decomposition) the energy density profile $E_M[s](t, \omega)$ at this frequency ω [see (20)]. A simple thresholding on this profile will allow to detect the beginning and the end of the corresponding note. The algorithm then loops by considering the "next" most energetic harmonic atom skipping all the atoms that have been "marked" as belonging to some formerly detected notes. The algorithm stops when the only harmonic atoms left have small energy.

Let us describe each step of this algorithm more precisely. First, of course, the harmonic matching pursuit is performed on the considered signal. At the beginning of the detection algorithm none of the harmonic atoms are marked.

1) Locate the most energetic harmonic atom $\lambda_i h_i(t)$, which is not marked.
2) If $|\lambda_i^2|/||s||^2$ is smaller than a given threshold ϵ_{stop}, the algorithm stops.

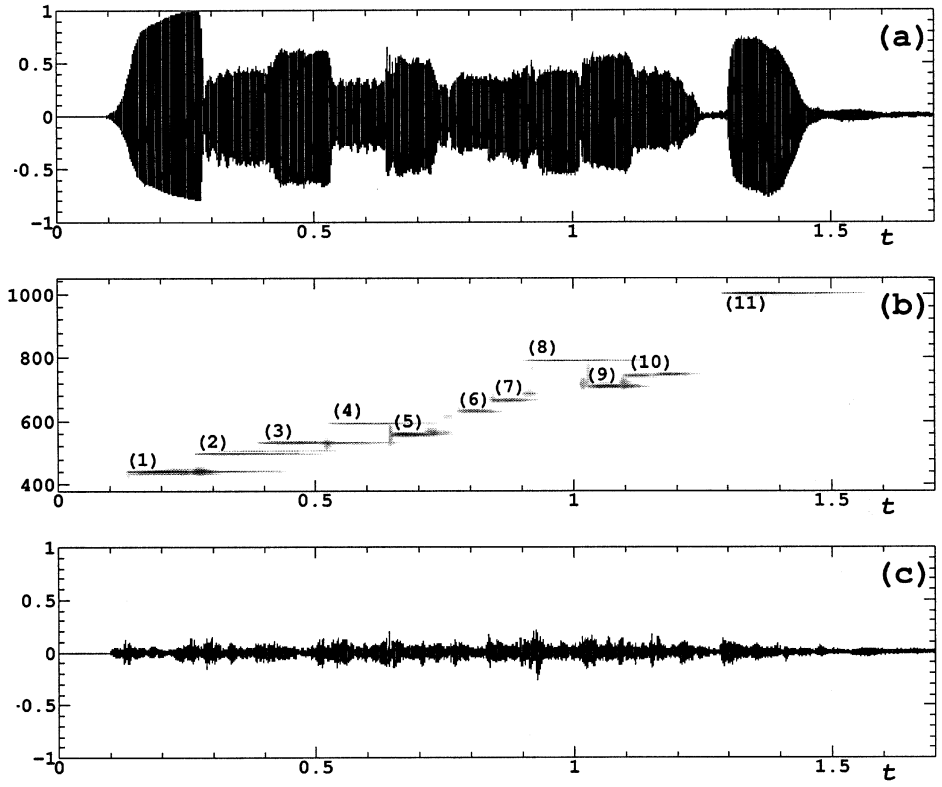

Fig. 3. Harmonic matching pursuit of a real audio signal. The analyzed signal corresponds to an excerpt of a recording of a solo clarinet piece. It consists in a melodic phrase made of 11 notes with very different durations and a lot of reverberation. The harmonic matching pursuit is performed using the asymmetric window $w(t)$ (cf. Fig. 2), with $M = 100, \xi_{\min} = 130$ Hz, and $\xi_{\max} = 1400$ Hz [see (5)]. (a) Graph of the audio signal $s(t)$. (b) Reduced energy density $E_M^{(r)}[s](t, \omega)$ [see (21)] obtained through the harmonic matching pursuit. It is gray-coded from the smallest values (white) to the largest values (black). To make this figure easier to read, each note has been indexed by order of appearance from the first one (1) to the last one (11). (c) Residual $R_M(t) = s(t) - s_M(t)$, where $s_M(t)$ is the resynthesized signal [see (22)]. Relative L^2 error is of the order of 9 dB.

3) The most energetic partial of this harmonic atom is considered to be the fundamental frequency $\omega = \xi_{i,k^*}$ of the note, i.e., $k^* = \arg\max_k |c_{i,k}|$.
4) Compute the energy density profile $D_\omega(t) = E_M[s](t, \omega)$ at this frequency.
5) Let u_i be the time location of the considered harmonic atom $(s_i, u_i, \xi_{i,1}, \ldots, \xi_{i,K})$. Let $D_\omega^{(\max)}$ the maximum value of $D_\omega(t)$ in a neighborhood of u_i (of size of the order of s_i). Compute the largest (resp. smallest) time $t < u_i$ (resp. $t > u_i$) for which $10\log_{10}(D_\omega(t)/D_\omega^{(\max)})$ is larger than a given threshold θ_{beg} (resp. θ_{end}). The so-obtained time location t_{beg} (resp. t_{end}) is considered as being the beginning (resp. end) of the note.
6) If the duration $t_{end} - t_{beg}$ is large enough (i.e., $> \Delta t_{\min}$), a note is detected at frequency ω at time t_{beg} till time t_{end}.
7) We mark the current harmonic atom along with all the harmonic atoms $h_j(t)$ that correspond to the same note, i.e., which satisfy $t_{beg} \leq u_j \leq t_{end}$ and for which at least one partial atom $c_{j,k} g_{j,k}$ satisfies, for a given threshold ϵ_{mark}

$$\max_t WV[g_{j,k}](t, \omega) > \epsilon_{\text{mark}}.$$

8) Go back to step 1.

B. Note Detection With Some Musical Signals

In this section, we apply the note detection algorithm described in the previous section to some musical signals. The parameters for the algorithm have been chosen in the following way: $\epsilon_{\text{stop}} = 0.01$, $\theta_{\text{beg}} = \theta_{\text{end}} = -14$ dB, $\Delta t_{\min} = 0.03$ s and $\epsilon_{\text{mark}} = 2.10^{-5}$.

We first apply it to the clarinet signal previously analyzed (cf. Fig. 3). As illustrated in Fig. 4(a), all the notes are successfully detected, although they are of very different durations and, as shown by the time-frequency representation, some of them overlap each other (due to reverberation). Moreover, the beginning of each note is very accurately estimated. This estimation (i.e., step 5 of the algorithm) is illustrated on Fig. 4(b). Let us point out that the detection of notes of very different durations and of their starting time locations using "standard" techniques is difficult. In the particular case where the musical instrument that is playing is a piano, it has been shown in [25] that, if one performs an extensive learning phase on the specific piano that is used, then STFT-based algorithms can achieve polyphonic note detection with few errors. However, in the case there is none or little prior information on the specific instruments that are playing, these algorithms no longer apply and precise note detection using STFT techniques becomes really difficult. A major problem one has to face when using STFT is the fact that one has to choose, once for all, the size of the window. Ideally, one

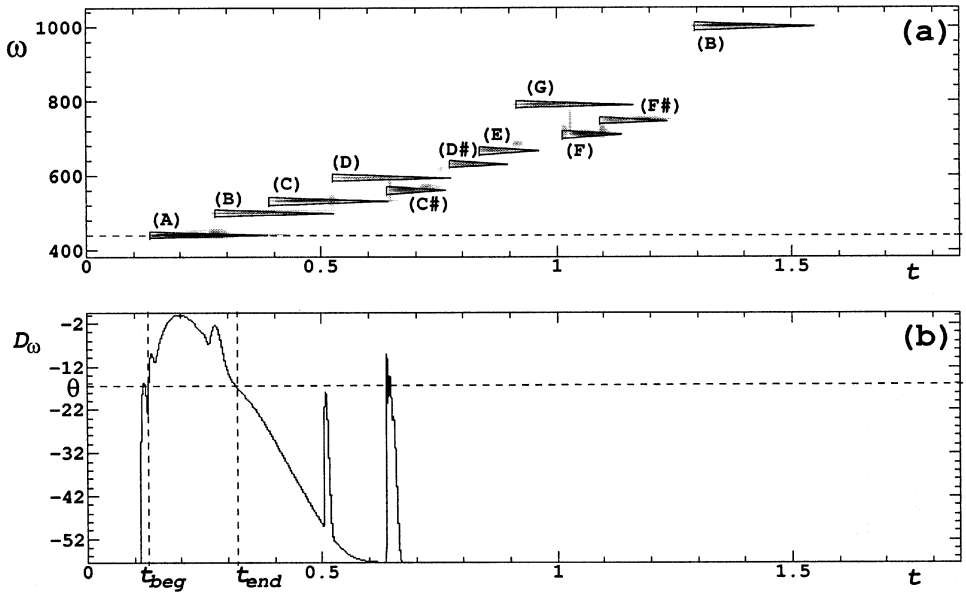

Fig. 4. Note detection on a clarinet signal. The signal is the same as the one displayed in Fig. 3(a). (a) Reduced energy density $E_M^{(r)}[s](t, \omega)$ defined by (21). The harmonic matching pursuit is performed using the same parameters as for Fig. 3. The detected notes obtained using the detection algorithm described in Section VIII-A are indicated using triangle shapes and the corresponding symbolic names for each note is in parentheses (using the standard notation A, B, C, D, E, F, G). (b) Function $D_\omega(t) = E_M^{(r)}[s](t, \omega)$ for the first note (A). The beginning time $t_{\rm beg}$ and ending time $t_{\rm end}$ of this note are estimated using simple thresholding (i.e., step 5 of the algorithm) with $\theta = \theta_{\rm beg} = \theta_{\rm end} = 19$ dB.

would want to use both a very short window (for very short notes or for estimating accurately the starting time of each note) and longer windows (of the order of the note durations). In a way, the harmonic matching pursuit changes the size of the window adaptatively according to these requirements and makes the detection very accurate.

The next signal we have tested our note detection algorithm on is a recording of the beginning of the *Chromatic Fantasy* by Bach [2]. Detection is much harder than on the previous signal: This recording involves much more reverberation, the melody is really fast, and the attacks of the notes are very soft. Moreover, since the melody is made of three successive scales (ascending, descending, and then ascending again), there are a lot of reverberating octaves or fifths, which make the detection all the more difficult. Actually, at a given time t, the only way to "understand" that an octave has been played (and not just a single note) is to look at the energy density at time locations that can be arbitrarily far from time t, i.e., at times when only one of the two notes of the octave could be heard. Although this is very hard to achieve using a regular algorithm based on a STFT, this is automatically performed when using the harmonic matching pursuit. As illustrated in Fig. 5, the detection is quite good. The reverberating fifths and the octaves are all detected. Moreover, although the attack of each note is very soft, all the notes are detected.

However, the algorithm makes a few mistakes. Most mistakes are due to the sensitivity of the algorithm to the parameters values (mainly $\theta_{\rm beg}$ and $\theta_{\rm end}$). The value of $\theta_{\rm beg}$ must be chosen relatively to how hard the attacks of the notes are. As seen in Fig. 5, most of the attack time locations are well estimated except for very few (e.g., the "A" detected at time $t \simeq 0.8$, which has been detected to start after the "B" leading to an inversion of the scale). Moreover, low values for $\theta_{\rm beg}$ and $\theta_{\rm end}$ will lead the algorithm to merge notes that have the same frequency and that are close enough one to each other, whereas large values will lead the algorithm to split single notes into two notes, as if the note was played twice. Actually, in Fig. 5, one can see that all the notes are detected, but the "G" (starting at time $t \simeq 2.2$) is detected twice (one after the other) though only one "G" has been played.

Let us point out that these decisions (mainly the one note versus two notes decision and the estimation of the time location of the beginning of a note) are really hard to make, especially because of the reverberation and of the soft attacks. If the attacks were a little harder, just decreasing the θ_{beg} would improve the result quite a bit.

Considering the difficulty of the note detection on this signal, the fact that this rather simple algorithm succeeds in finding all the notes and in estimating precisely most of their starting time locations, makes, we think, the harmonic matching pursuit a very promising tool for note detection.

IX. Conclusion

The flexibility of the matching pursuit paradigm makes it possible to design dictionaries of elementary waveforms that reflect the expected structures of the analyzed signals. Harmonic structures, which are common in audio signals, are easily described as linear combinations of a few quasiorthogonal Gabor atoms. This enables the efficient realization of a harmonic matching pursuit decomposition. One can indeed notice that the complexity $\mathcal{O}(MKN)$ of the fast harmonic matching pursuit is essentially that of building the approximant $\sum_{m,k} c_{m,k} g_{m,k}$, i.e., the cost of selecting the harmonic atoms of interest is negligible.

Because of its demonstrated ability to decompose a musical recording into harmonic structures of very different durations and that could overlap each other, the harmonic matching pursuit is a

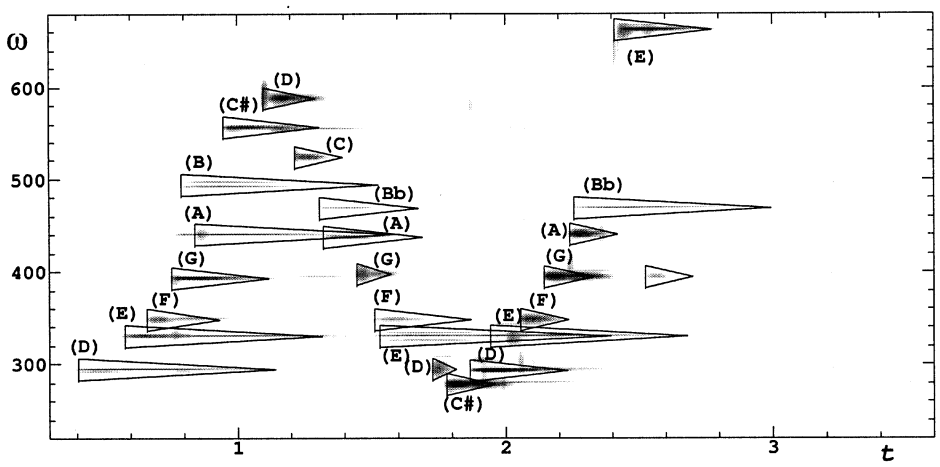

Fig. 5. Note detection on a piano signal. The signal corresponds to the beginning of the *Chromatic Fantasy* by Bach. It is basically made of three successive scales (ascending, descending, and then ascending again) played really fast with very soft attacks and a lot of reverberation. (a) Reduced energy density $E_M^{(r)}[s](t,\,\omega)$ as defined in (21). The harmonic matching pursuit is performed using the asymmetric window $w(t)$ (cf. Fig. 2), with $M = 100$, $\xi_{\min} = 130$ Hz and $\xi_{\max} = 1700$ Hz [see (5)]. The detected notes obtained using the detection algorithm described in Section VIII-A are indicated using triangle shapes, and the corresponding symbolic names for each note are in parentheses (using the standard notation A, B, C, D, E, F, G).

very promising tool for note detection. However, the note detection algorithm we proposed is still too sensitive to its parameter values. We actually believe that this sensitivity is inherent to the use of the harmonic matching pursuit itself rather than to the detection algorithm and that, consequently, the pursuit should not be used "as is." We believe that the selection of harmonic atoms for note detection should rather be done simultaneously with the detection algorithm itself, which may imply using tracking techniques [12], [13] and penalizing those harmonic atoms where the coefficients $\{c_{m,\,k}\}_{k=1}^{K}$ "oscillate" to much.

As a last remark, we believe the "subspace matching pursuit" framework that we have defined in this paper may be the basis for other applications where it is possible to use dictionaries that are the union of the unit spheres of small dimensional spaces spanned by simple quasiorthogonal atoms.

Appendix A
Proof of Theorem 1

To prove this theorem, we are going to use [18, Th. 2.1], which states the condition of convergence for the *approximate weak greedy algorithms*. This theorem shows that in order to prove the convergence of our modified pursuit, one just needs to prove the two following points.

1) For some sequence $0 \leq \alpha_m \leq 1$, which satisfies $\sum_m \alpha_m/m = +\infty$, h_m satisfies

$$|\langle R_{m-1},\, h_m\rangle| \geq \alpha_m \sup_{h\in\mathcal{D}_h} |\langle R_{m-1},\, h\rangle|\,. \tag{24}$$

2) For some $0 < \delta \leq 1$, the coefficient λ_m is *approximately* $\langle R_{m-1},\, h_m\rangle$:

$$\lambda_m = (1+\epsilon_m)\,\langle R_{m-1},\, h_m\rangle \tag{25}$$

with $\epsilon_m \in [-1+\delta,\, 1-\delta]$.

(Let us note that the condition for convergence in [18] is slightly weaker.)

For any $\gamma = (s,\, u,\, \xi_1,\, \ldots,\, \xi_K)$, let $G_\gamma = (\langle g_k,\, g_l\rangle)_{1\leq k,\, l\leq K}$, the Grammian matrix of the family $\{g_k = g_{s,\,u,\,\xi_k},\, 1 \leq k \leq K\}$. One can easily check that (6) implies that the eigenvalues of G_γ lie within $[\delta,\, 2-\delta]$. Hence, $\{g_k,\, 1 \leq k \leq K\}$ is a linearly independent family of vectors, as well as its biorthogonal basis $\{\tilde{g}_k,\, 1 \leq k \leq K\}$ (characterized by $\langle \tilde{g}_k,\, g_l\rangle = \delta_{kl}$) in $\mathcal{V}_\gamma$. As a result, the extremal eigenvalues of the positive definite quadratic form

$$v \mapsto Q_\gamma(v) := \sum_{k=1}^{K} |\langle v,\, g_l\rangle|^2$$

(restricted to the finite dimensional subspace $\mathcal{V}_\gamma$) are equal to the extremal eigenvalues of G_γ. Hence, for all $v \in \mathcal{H}$ and all $\gamma \in \Gamma$

$$\delta \left\|P_{\mathcal{V}_\gamma} v\right\|^2 \leq Q_\gamma(P_{\mathcal{V}_\gamma} v) = Q_\gamma(v) \leq (2-\delta) \left\|P_{\mathcal{V}_\gamma} v\right\|^2. \tag{26}$$

Similarly for the dual basis, for all $v \in \mathcal{H}$ and all $\gamma \in \Gamma$, $(2-\delta)^{-1} \|P_{\mathcal{V}_\gamma} v\|^2 \leq \sum_k |\langle v,\, \tilde{g}_k\rangle|^2 \leq \delta^{-1} \|P_{\mathcal{V}_\gamma} v\|^2$. Thus, since h_m belongs to $\mathcal{V}_{\gamma_m}$ and $\|h_m\|^2 = 1$, one can show that there exists $\epsilon_m \in [-1+\delta,\, 1-\delta]$ such that $\sum_{k=1}^{K} |\langle h_m,\, \tilde{g}_{m,\,k}\rangle|^2 = (1+\epsilon_m)^{-1}$. Thus, one easily gets

$$\begin{aligned}\lambda_m &= \lambda_m(1+\epsilon_m) \sum_{k=1}^{K} |\langle h_m,\, \tilde{g}_{m,\,k}\rangle|^2 \\ &= (1+\epsilon_m) \sum_{k=1}^{K} \langle \lambda_m h_m,\, \tilde{g}_{m,\,k}\rangle \langle \tilde{g}_{m,\,k},\, h_m\rangle\,.\end{aligned}$$

Since, from (11), one gets that $\langle \lambda_m h_m,\, \tilde{g}_{m,k}\rangle = \langle R_{m-1},\, g_{m,k}\rangle$, we finally get

$$\lambda_m/(1+\epsilon_m) = \sum_{k=1}^{K} \langle R_{m-1},\, g_{m,\,k}\rangle \langle \tilde{g}_{m,\,k},\, h_m\rangle = \langle R_{m-1},\, h_m\rangle.$$

Hence, this proves point 2) [i.e., (25)].

From (11), one gets

$$Q_{\gamma_m}(R_{m-1}) = \sum_{k=1}^{K} |\langle R_{m-1},\, g_{m,k}\rangle|^2 = \langle R_{m-1},\, \lambda_m h_m\rangle\,.$$

Then, using (25)

$$Q_{\gamma_m}(R_{m-1}) = (1+\epsilon_m)\,|\langle R_{m-1},\, h_m\rangle|^2\,. \tag{27}$$

On the other hand, (9) and (26), along with the fact that $\epsilon_m \in [-1+\delta,\, 1-\delta]$, give

$$\begin{aligned}\frac{Q_{\gamma_m}(R_{m-1})}{1+\epsilon_m} &\geq \frac{\rho_m^2}{1+\epsilon_m}\,\sup_{\gamma\in\Gamma} Q_\gamma(R_{m-1})\\ &\geq \rho_m^2 \frac{\delta}{2-\delta}\,\sup_{\gamma\in\Gamma}\left\|P_{\mathcal{V}_\gamma} R_{m-1}\right\|^2.\end{aligned}$$

Combining this last equation with (27), we get

$$|\langle R_{m-1},\, h_m\rangle| \geq \sqrt{\frac{\delta}{2-\delta}}\;\rho_m\,\sup_{h\in\mathcal{D}_h} |\langle R_{m-1},\, h\rangle|\,.$$

Since Theorem 1 assumes that $\sum_m \rho_m/m = +\infty$, this last equation proves point 1) [i.e., (24) with $\alpha_m = \sqrt{\delta/(2-\delta)}\,\rho_m$].

In order to apply [18, Th. 2.1], we just need to check that $\mathcal{D}_h$ is complete, but this is easily done because it contains $\mathcal{D}_g$, which is complete as soon as the window is smooth and satisfies $|w(t)| = \mathcal{O}((1+t^2)^{-1})$ [22].

Appendix B
Proof of (14)

From (11), one gets

$$\|R_{m-1}\|^2 - \|R_m\|^2 = \left(1-\epsilon_m^2\right)|\langle R_{m-1},\, h_m\rangle|^2\,. \tag{28}$$

Using (25), this last equation becomes

$$\|R_{m-1}\|^2 - \|R_m\|^2 = \frac{1-\epsilon_m}{1+\epsilon_m}\,\lambda_m^2. \tag{29}$$

On the other hand, from (27), one deduces that

$$\begin{aligned}\left(1-\epsilon_m^2\right)|\langle R_{m-1},\, h_m\rangle|^2 &= (1-\epsilon_m) Q_{\gamma_m}(R_{m-1})\\ &= (1-\epsilon_m)\sum_{k=1}^{K} |c_{m,k}|^2.\end{aligned}$$

Using (28), we get

$$\|R_{m-1}\|^2 - \|R_m\|^2 = (1-\epsilon_m)\sum_{k=1}^{K} |c_{m,k}|^2. \tag{30}$$

Hence, using (29) and (30)

$$\begin{aligned}\|s\|^2 &= \sum_{m=1}^{M} \frac{1-\epsilon_m}{1+\epsilon_m}\,|\lambda_m|^2 + \|R_M\|^2\\ &= \sum_{m=1}^{M}\sum_{k=1}^{K}(1-\epsilon_m)|c_{m,k}|^2 + \|R_M\|^2\,.\end{aligned}$$

Since $|\epsilon_m| \leq 1-\delta$, if we assume that $\delta \simeq 1$, we obtain the desired approximations.

Acknowledgment

The authors would like to thank S. Mallat, from Ecole Polytechnique, for his interest and his suggestions in this research, and J. Abadía Domínguez for technical help. The clarinet sound recording was kindly provided by X. Rodet, from IRCAM, whom they also wish to thank for interesting discussions. All the numerical computations and figures were obtained using LastWave [3], a freely available software under the GPL license.

References

[1] J. Abadía Domínguez, "Análisis de senales sonoras mediante técnicas de Seguimiento Adaptativo. Reconocimiento automático de música," CMAP, Ecole Polytechnique, Tech. Rep., 2000.
[2] J. S. Bach, *Chromatic Fantasy and Fugue by A. Brendel (piano)*. Eindhoven, The Netherlands: CD Philips CD, 442 400-2.
[3] E. Bacry. LastWave software (GPL license). [Online]. Available: http://wave.cmap.polytechnique.fr/soft/LastWave/.
[4] F. Bergeaud, "Représentations adaptatives d'images numériques, Matching Pursuit," Ph.D. dissertation, Ecole Centrale Paris, Paris, France, 1995.
[5] P. Boulez, "Dialogue de l'ombre double," in "*Pierre Boulez*": Erato Disques, CD, 1991, 2292-45 648-2.
[6] S. Chen and D. L. Donoho, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, Jan. 1999.
[7] B. Doval, "Estimation de la fréquence fondamentale des signaux sonores," Ph.D. dissertation, Univ. Paris VI, Paris, France, 1994.
[8] B. Doval and X. Rodet, "Fundamental frequency estimation and tracking using maximum likelihood harmonic matching and HMMs," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, Minneapolis, MN, April 1993, pp. 221–224.
[9] P. Flandrin, *Temps-Fréquence* Hermes, Paris, France, 1993.
[10] H. Fletcher, "Normal vibration frequencies of a stiff piano string," *J. Acoust. Soc. Amer.*, vol. 36, no. 1, pp. 203–209, 1962.
[11] J. H. Friedman and W. Stuetzle, "Projection pursuit regression," *J. Amer. Stat. Assoc.*, vol. 76, pp. 817–823, 1981.
[12] G. García, P. Depalle, and X. Rodet, "Tracking of partial for additive sound synthesis using hidden Markov models," in *Proc. Int. Comput. Music Conf.*, Tokyo, Japan, 1993, pp. 94–97.
[13] D. Geman and B. Jedynak, "An active testing model for tracking roads in satellite images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 1–14, Jan. 1996.
[14] M. Goodwin, "Matching pursuit with damped sinusoids," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, 1997.
[15] M. Goodwin and M. Vetterli, "Matching pursuit and atomic signal models based on recursive filter banks," *IEEE Trans. Signal Processing*, vol. 47, pp. 1890–1902, July 1999.
[16] R. Gribonval, "Approximations nonlinéaires pour l'analyse de signaux sonores," Ph.D. dissertation, Univ. Paris IX, Dauphine, France, Sept. 1999.
[17] R. Gribonval, P. Depalle, X. Rodet, E. Bacry, and S. Mallat, "Sound signals decomposition using a high resolution matching pursuit," in *Proc. Int. Computer Music Conf.*, Aug. 1996, pp. 293–296.
[18] R. Gribonval and M. Nielsen, "Approximate weak greedy algorithms," *Adv. Comput. Math.*, vol. 14, no. 4, pp. 361–378, May 2001.
[19] P. J. Huber, "Projection pursuit," *Ann. Statist.*, vol. 13, no. 2, pp. 435–475, 1985.
[20] S. Jaggi, W. C. Carl, S. Mallat, and A. S. Willsky, "High resolution pursuit for feature extraction," *J. Appl. Comput. Harmon. Anal.*, vol. 5, no. 4, pp. 428–449, Oct. 1998.
[21] L. K. Jones, "On a conjecture of Huber concerning the convergence of PP-regression," *Ann. Statist.*, vol. 15, pp. 880–882, 1987.
[22] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, pp. 3397–3415, Dec. 1993.
[23] S. Qian and D. Chen, "Signal representation using adaptive normalized Gaussian functions," *Signal Process.*, vol. 36, no. 1, pp. 1–11, 1994.
[24] X. Rodet, "Time-domain formant-wave functions synthesis," in *Spoken Language Generation and Understanding C: Mathematical and Physical Sciences*, J. C. Simon, Ed. New York: D. Reidel, 1980, ch. 4—Speech Synthesis, pp. 429–441.

[25] L. Rossi, "Identification de sons polyphoniques de piano," Ph.D. dissertation, Univ. Corse, Corse, France, Jan. 1998.
[26] B. Torrésani, "Wavelets associated with representations of the affine Weyl–Heisenberg group," *J. Math. Phys.*, vol. 32, pp. 1273–1279, May 1991.
[27] G. H. Watson and K. Gilholm, "Signal and image feature extraction from local maxima of generalized correlation," *Pattern Recogn.*, vol. 31, no. 11, pp. 1733–1745, 1998.

Rémi Gribonval graduated from École Normale Supérieure, Ulm, Paris, France, in 1997 and the Ph.D. degree in applied mathematics from the Université Paris-IX, Dauphine, Paris, France, in 1999.

Since 2000, he is a Research Associate with the French National Center for Computer Science and Control (INRIA), IRISA, Rennes, France. His current research interests are in adaptive techniques for the representation and classification of audio signals with redundant systems.

Emmanuel Bacry graduated from École Normale Supérieure, Ulm, Paris, France, in 1990 and received the Ph.D. degree in applied mathematics from the University of Paris VII, Paris, France, in 1992. He rceived the "habilitation à diriger des recherches" from the same university in 1996.

Since 1992, he has been a researcher at the Centre National de Recherche Scientifique (CNRS). After spending four years with the applied mathematics department of Jussieu (Paris VII), he moved to the Centre de Mathématiques Appliquées (CMAP), École Polytechnique, Paris, France, in 1996. During the same year, he became a part-time assistant professor at Ecole Polytechnique. His research interests include signal processing, wavelet transforms, fractal and multifractal theory with applications to finance and various domains such as sound processing.