ANALYSIS OF SOUND SIGNALS WITH HIGH RESOLUTION MATCHING PURSUIT

R. Gribonval E. Bacry S. Mallat† Ph. Depalle and X. Rodet‡*

IRCAM, 1 place Igor-Stravinsky, 75004 PARIS, FRANCE.
e-mail : gribonva@clipper.ens.fr, fax : (33)-1-42772947

ABSTRACT

Sound recordings include transients and sustained parts. Their analysis with a basis expansion is not rich enough to represent efficiently all such components. Pursuit algorithms choose the decomposition vectors depending upon the signal properties. The dictionary among which these vectors are selected is much larger than a basis. Matching Pursuit is fast to compute, but can provide coarse representations. Basis Pursuit gives a better representation but is very expensive in terms of calculation time. This paper develops a High Resolution Matching Pursuit : it is a fast, high time-resolution, time-frequency analysis algorithm, that makes it likely to be used for musical applications.

1. INTRODUCTION

The complexity of structures encountered in a sound recording requires to develop adaptive low-level representations. Although the decomposition of such a signal in a basis entirely characterizes it, this basis is a minimal set of vectors that is not rich enough to represent efficiently all components. Some signal structures are diffused across many basis elements and are thus difficult to analyze from this expansion.

Indeed, sound recordings include transients that are well represented by short waveforms, and sustained parts that are more efficiently decomposed over long waveforms with short frequency support. Pursuit algorithms choose the decomposition vectors depending upon the signal properties. These vectors are selected among a family of waveforms that is much larger than a basis, which is called a dictionary.

The High Resolution Matching Pursuit (HR-MP) developed in this paper is a fast algorithm providing high time-resolution time-frequency representations, that enables it to be used for musical applications.

2. MATCHING PURSUIT

A *dictionary* is a family of vectors $\mathcal{D} = (g_\gamma)_{\gamma\in\Gamma}$ included in a Hilbert space H, with a unit norm $\|g_\gamma\| = 1$. A *matching pursuit* is an iterative algorithm that decomposes the signal over dictionary vectors as follows.

Let $R^0 f = f$. We suppose that we have computed the n^{th} order residue $R^n f$, for $n \geq 0$. We then choose an element $g_{\gamma_n} \in \mathcal{D}$ which "closely" matches the residue $R^n f$, in the sense that

$$|C(R^n f, g_{\gamma_n})| = \sup_{\gamma\in\Gamma} |C(R^n f, g_\gamma)|, \qquad (1)$$

where $C(f, g_\gamma)$ is a *correlation function* that measures the similarity between f and g_γ.

The residue $R^n f$ is then sub-decomposed into

$$R^n f = C(R^n f, g_{\gamma_n}) g_{\gamma_n} + R^{n+1} f, \qquad (2)$$

which defines the residue at the order n+1. In the Matching Pursuit (MP), initially introduced by Mallat and Zhang [1], the correlation function that is used is $C(f, g_\gamma) = < f, g_\gamma >$. The error $\|R^n f\|$ is then proved to decay to zero. Thus by

---
*CNRS Paris 7
†CMAP Ecole Polytechnique
‡IRCAM

iterating Eq. (2) we obtain

$$f = \sum_{n=0}^{+\infty} C(R^n f, g_{\gamma_n}) g_{\gamma_n}. \tag{3}$$

The structure of MP enables it to be implemented with a fast algorithm.

3. GABOR DICTIONARY

To analyze time and frequency localization properties of one-dimensional signals, such as speech or music recordings, we use a large dictionary of *time-frequency atoms.*

Let $g(t) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{t^2}{2\sigma^2}\right)$ be a Gaussian function of unit norm. For any scale $s > 0$, modulation frequency ξ and translation u, we denote $\gamma = (s, u, \xi)$ and define

$$g_\gamma(t) = \frac{1}{\sqrt{s}} g\left(\frac{t-u}{s}\right) e^{i\xi t}. \tag{4}$$

The index γ is an element of the set $\Gamma = R^+ \times R^2$. The function $g_\gamma(t)$ is centered at the abscissa u and its energy is concentrated in a neighborhood of u, whose size is proportional to s. Its Fourier transform is centered at the frequency $\omega = \xi$, and its energy is concentrated in a neighborhood of ξ, whose size is proportional to $1/s$. The time-frequency energy distribution of $f(t)$ is then defined by

$$Ef(t,\omega) = \sum_{n=0}^{+\infty} |C(R^n f, g_{\gamma_n})|^2 W g_{\gamma_n}(t,\omega). \tag{5}$$

where $Wg_{\gamma_n}(t,\omega)$ is the Wigner distribution of g_{γ_n}, *i.e.* a two-dimensional Gaussian "blob" in the time-frequency plane. Figure 1 and 2 display such time-frequency energy distributions.

4. HIGH RESOLUTION MATCHING PURSUIT

The Matching Pursuit is a greedy algorithm in that it optimizes at each step the amount of the signal energy it grasps. This often leads to a choice of features which globally fits the signal structures but is not best adapted to its local structures. Indeed, for instance, a signal composed of two bumps modulated by a sinusoidal wave at frequency ξ (Figure 1-a) is first decomposed into a large atom at frequency ξ (middle horizontal line on Figure 1-a-MP) that covers the time support of both bumps. Then, in order to remove the energy created between the two bumps by this first atom, MP chooses two atoms of the same size as the first one, with frequencies $\xi + \Delta\xi$ (upper line) and $\xi - \Delta\xi$ (lower line).

Aiming at avoiding this problem, Donoho and Chen [2] introduced the Basis Pursuit, which makes a full optimization, by minimizing $\sum_{\gamma\in\Gamma} |\alpha_\gamma|$ over all possible decompositions $f = \sum_{\gamma\in\Gamma} \alpha_\gamma g_\gamma$. However this leads to large scale linear-programming problems and therefore is very expensive in terms of calculation time.

The new algorithm, that we called High Resolution Matching Pursuit (HRMP), is an enhanced version of Matching Pursuit (MP), extending to time-frequency dictionaries the pursuit over non-modulated spline dictionaries introduced by Jaggi et. al. [3]. It uses a different correlation function, that allows the pursuit to emphasize local fit over global fit at each step. The fast algorithm structure of MP is however kept.

For each time-frequency atom g_γ a set I_γ of *sub-atom* indexes is introduced. I_γ corresponds to smaller atoms $g_{\gamma_i}, \gamma_i \in I_\gamma$ with a time support included in the support of g_γ, and modulated at the same frequency. Let suppose that the atom g_γ is chosen in a pursuit. $Rf = f - C(f, g_\gamma)g_\gamma$ becomes the residue of this pursuit on the signal f. For all $\gamma_i \in I_\gamma$, $< Rf, g_{\gamma_i} >$ represents the amount of "energy" of Rf located on the time-frequency support of g_{γ_i}. This amount must be smaller than the signal "energy" $< f, g_{\gamma_i} >$ at the same location. Moreover the corresponding decrease $< C(f, g_\gamma)g_\gamma, g_{\gamma_i} >$ of signal energy cannot be greater than the initial signal energy itself. This is formalized in Equations 6 and 7 :

$$|< Rf, g_{\gamma_i} >| \leq |< f, g_{\gamma_i} >|, \tag{6}$$

$$|< C(f, g_\gamma)g_\gamma, g_{\gamma_i} >| \leq |< f, g_{\gamma_i} >|. \tag{7}$$

From these relations, we derive the new correlation function $C(f, g_\gamma)$, which maximizes the amount of signal energy that the pursuit can grasp, when

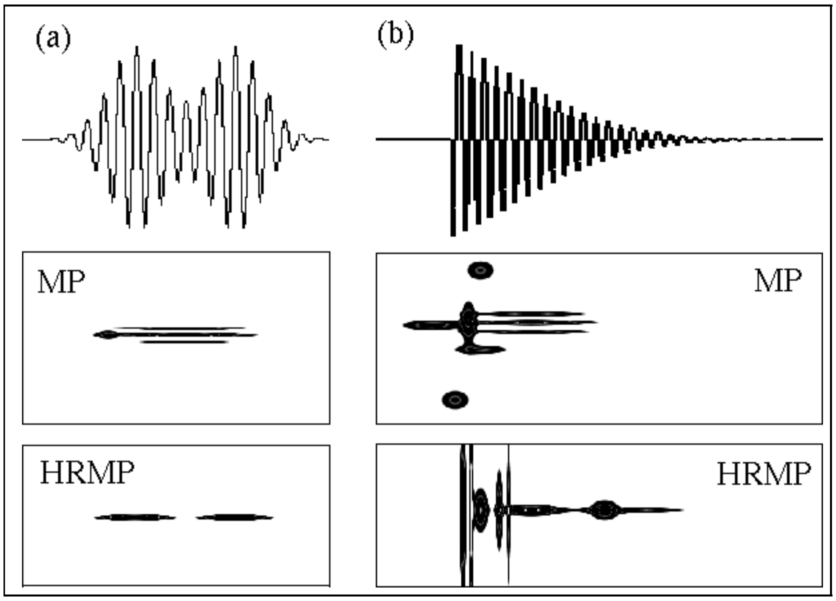

Figure 1: Time-Frequency distributions of signals (top) obtained with MP (middle) and HRMP (bottom):
(a) two close bumps, with a four atom decomposition
(b) an attack pattern, with a ten atom decomposition

choosing the atom g_γ :

$$C(f, g_\gamma) = \varepsilon \min_{\gamma_i \in I_\gamma} \frac{|< f, g_{\gamma_i} >|}{|< g_\gamma, g_{\gamma_i} >|} \tag{8}$$

where ε is evaluated as follows:

- if $< f, g_{\gamma_i} >$ have the same sign, for all $\gamma_i \in I_\gamma$, then ε is this common sign.
- else $\varepsilon = 0$.

In MP, the inner-product, used as a correlation function between a time-frequency atom and an audio signal, disregards whether the signal contains energy on the whole time-frequency support of the chosen atom. On the contrary, the new correlation function avoids creating energy at time locations where there was none. It can thus distinguish close time features as shown in Figure 1-a-HRMP. Moreover it can avoid pre-echo effects, *i.e.* creation of energy just before the beginning of the sound. Indeed, as shown in Figure 1-b, MP introduces a pre-echo effect by choosing atoms that overlap the attack time-location, whereas HRMP does not choose any such atom.

Figure 2 displays the time-frequency distribution of a piano note obtained from a HRMP analysis. It represents simultaneously structures of very different scales. First, horizontal lines, corresponding to large scale atoms, describe the harmonic structures : the bottom line corresponds to the fundamental frequency of the piano note at 830 Hz, and the lines above display the partials ; in the bottom left part, shorter horizontal lines display the resonance of the piano's sounding board at 20 Hz. Then vertical features, corresponding to fine scale transitory structures, describe both the attack, at the beginning of the note (left part), and the fall back of the piano's damper on the string, at its end (bottom right part).

Because of the new correlation function, the atoms chosen for the decomposition have a smaller time support than with a usual Matching Pursuit decomposition, hence, because of Heisenberg inequalities, they also have a larger frequency support. HRMP thus performs a higher time-resolution decomposition than MP, but its frequency-resolution is decreased. However for audio applications the most important is to keep a good lo-

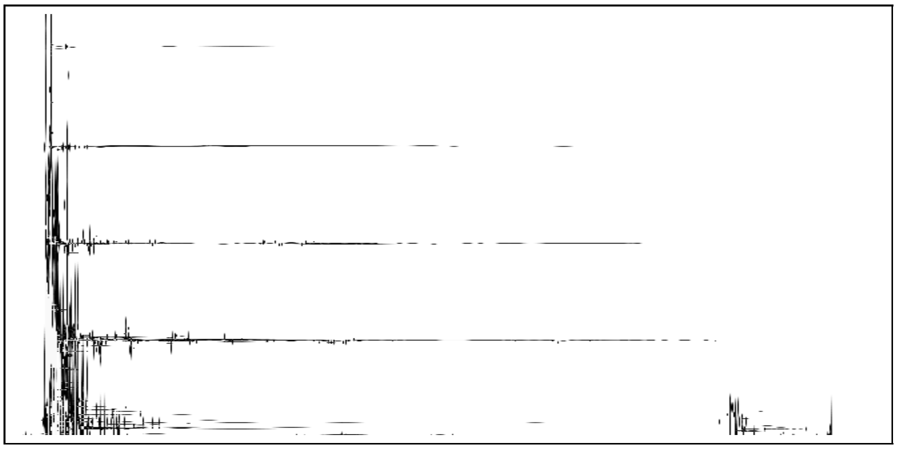

Figure 2: Time-Frequency distribution of a piano note with HRMP

calization of the attacks, because the ear is very sensitive to transients : hearing the attack of a musical instrument is almost sufficient to identify it. HRMP is thus adapted to the requirements of sound recording analysis. It is likely to be adapted to many musical applications, for example to allow the extraction of the parameters of formant-wave functions synthesizers [4], which could not be achieved with existing analysis processes.

5. REFERENCES

[1] S. Mallat and Z. Zhang. Matching pursuit with time-frequency dictionaries. *IEEE Trans. Signal Process.*, 41(12):3397–3415, December 1993.

[2] S. Chen and D.L. Donoho. Atomic decomposition by basis pursuit. Technical report, Statistics Department, Stanford University, 1995.

[3] S. Jaggi, W.C. Carl, S. Mallat, and A.S. Willsky. High resolution pursuit for feature extraction. Technical report, MIT, November 1995.

[4] X. Rodet. Time-domain formant-wave functions synthesis. In *Spoken Language Generation and Understanding*, pages 429–441. D. Reidel Publishing Company, 1980.