Sound Signals Decomposition Using a High
Resolution Matching Pursuit

R. Griboval* Ph. Depalle†
IRCAM, Analysis/Synthesis Department
1 place Igor-Strausssky,
F-75004 PARIS, FRANCE
fax: (33)-1-47772947

X. Rodet** E. Bacry†† S. Mallat†††
CMAP,
Ecole Polytechnique
F-91120 PALAISEAU CEDEX, FRANCE

Abstract
Sound recordings include transients and sustained parts. Their analysis with a ba-
sis expansion is not rich enough to repres-
ent efficiently all such components. Pur-
suit algorithms choose the decomposition vectors depending upon the signal pro-
erties. 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. Ba-
sis Pursuit gives a better representation but is very expensive in terms of calculation time. This paper develops a High Resolu-
tion Matching Pursuit: it is a fast, high
time-resolution, time-frequency analysis al-
gorithm, that makes it likely to be used for musical applications.

1 Introduction
The complexity of structures encountered in sound signals requires the development of adaptive low-
level representations in order to provide mean-
ingful analysis. Usual time-frequency analysis
methods, such as Wavelet [KM88] or Short Time Fourier Transform [6787] [Moo76], perform a de-
composition of the signal according to a given
fixed basis. Although such a decomposition en-
tirely characterizes the signal, a basis is a minimal set of vectors that is not rich enough to repre-
sent efficiently all components. Indeed, some sig-
nal structures may be diffused across many basis
elements: such an expansion could then become
difficult to correlate with perceptual entities.

Actually, sound signals include transients that are well represented by short waveforms, and sus-

*IRCAM, griboval@clipper.ccv.fr
†IRCAM, phd@ircam.fr
‡IRCAM, rod@ircam.fr
††CMAP, bacry@cmap.polytechnique.fr
†††CMAP, mallat@cmap.polytechnique.fr

sained parts that are not efficiently decomposed
over long waveforms with short frequency support.

Lately, new adaptive approaches have been de-
developed in order to choose the decomposition
vectors depending upon the signal properties:
for example, Coifman and Wickerhauser [CW92]
[BCG94] introduced an adaptive "best basis" se-
lection, but such an analysis algorithm cannot dis-
tinguish overlapping features, such as a "click"
and a slide wave together, as far as it chooses an
orthogonal basis (among a given family of such
bases) to decompose the sound.

Pursuit algorithms, such as Matching Pursuit
(MP) [M93] or Basis Pursuit (BP) [CD95]
were designed to overcome these problems. The de-
composition vectors are selected among a redu-
dant family of elementary waveforms both well-
localized in time and frequency. This family of
time-frequency atoms, which is much larger than
a basis, is called a dictionary.

BP is fast to compute, but can provide coarse
representations. MP gives a better representa-
tion but is very expensive in terms of calcu-
lation time. The High Resolution Matching Pur-
suit (HRMP) developed in this paper (see also
[GRM96]) is a fast pursuit algorithm providing
high time-resolution time-frequency represen-
tations. It is designed to comply with a good time-
localization constraint, thus eliminating pre-echo
effects that MP introduced. Therefore it is likely to
give much better results for musical applica-
tions.

2 Matching Pursuits
A dictionary is a family of vectors \( D = \{ v_n \}_{n \geq 0} \)

includes in a Hilbert space \( H \), with a unit norm, \( \| v_n \| = 1 \). A matching pursuit is an iterative algo-
rithm that decomposes the signal over dictionary
vectors as follows.

Let \( D f = f \). We suppose that we have com-
puted the \( n\text{th} \) order residue \( R_n f \), for \( n \geq 0 \). We
then choose \( n \text{th} \) element \( v_n \in D \) which "cloudy"
matches the residue \( R^n f \), in the sense that

\[
|C(f, g_\nu) | = \sup_{\nu \in \mathbb{N}} |C(f, g_\nu)|.
\]

(1)

where \( C(f, g_\nu) \) is a correlation function that measures the similarity between \( f \) and \( g_\nu \).

The residue \( R^n f \) is then self-decomposed into

\[
R^n f = C(R^n f, g_\nu) g_\nu + R^{n+1} f,
\]

(2)

which defines the residue at the order \( n+1 \).

In the Matching Pursuit (MP), initially introduced by Mallat and Zhang [MZ95], the correlation function that is used is the inner product \( C(f, g_\nu) \equiv \langle f, g_\nu \rangle \). Other correlation functions can be used: indeed, the object of this paper is to introduce a correlation function that is adapted to the requirements of audio signals representation.

This correlation function will be given by Eq. (6) and will lead to our new pursuit algorithm, High Resolution Matching Pursuit (HRMP).

With either of those correlation functions, the energy of the error \( |R^n f|^2 \) is proved to decay to zero. Thus by iterating Eq. (2) we obtain the atomic decompositions of the signal

\[
f = \sum_{\nu \in \mathbb{N}} C(R^n f, g_\nu) g_\nu.
\]

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

\[
g_\gamma(t) = \frac{1}{\sqrt{s}} \left(\frac{t-u}{s}\right)^{\gamma^T} g(t).
\]

(4)

The index \( \gamma \) is an element of the set \( \Gamma = \mathbb{R}^+ \times \mathbb{R}^2 \).

The function \( g_\gamma(t) \) is centered at the absciss \( u \) 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 \( \xi \), whose size is proportional to \( 1/s \).

Short scale atoms almost correspond to "clicks", whereas large scale atoms are nearly pure sine waves. This dictionary is thus likely to comply with the representation of transient structures as well as of stationary features.

4 Energy Distributions

The time-frequency energy distribution of \( f(t) \) is then defined by

\[
E(f, \omega) = \sum_{n \in \mathbb{N}} |C(R^n f, g_\nu)|^2 |W_{g_\nu} (t, \omega)|^2
\]

(5)

where \( W_{g_\nu}(t, \omega) \) is the Wigner distribution of \( g_\nu \), \( \nu \rightarrow \omega \) a two-dimensional Gaussian "blob" in the time-frequency plane. Figures 1, 2-MP and 2-HRMP display such time-frequency energy distributions.

For example, when applied to a medium pitch (e.g., G sharp) piano sound, a Matching Pursuit provides a time-frequency representation (Figure 4) that displays simultaneously structures of very different scales. At first, one can see the quasi-harmonic structure of the note. It is displayed by horizontal lines, corresponding to large scale, well-localized in frequency atoms. Then, below 100 Hz, shorter horizontal lines display the partials of the quasi-harmonic resonance of the piano's sounding board, at a fundamental frequency of approximately 20 Hz. Lastly, vertical features, corresponding to fine-scale transient structures describe both the attack at the beginning of the note, and the fall back of the piano's damper on the string at its end.

5 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 grasp. 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 \( \xi \) (Figure 2-a) is first decomposed into a large atom at frequency \( \xi \) (middle horizontal line on Figure 2-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 \( \xi \pm \Delta \xi \) (upper line) and \( \xi - \Delta \xi \) (lower line).

Moreover, we observed that MP does not keep a good localization of attack patterns (Figure 2-b-MP), which leads to a little, but still audible, pre-echo at pre-synthesis stage. This is due to the atoms selection criteria that allows the creation of energy where there was none previously.

Aiming at avoiding this problem, Donoho and Chen [CD95] introduced the Basis Pursuit, which makes a full optimization, by minimizing

\[
\sum_{n \in \mathbb{N}} |s_n| \quad \text{over all possible decompositions}
\]

Gribowol et al.

ICMC Proceedings 1996

294
\[ f = \sum_{i=1}^{\infty} c_i g_i. \] 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. [JCMW98]. 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 \( g_i \), a set \( I_i \) of sub-atom indexes is introduced. \( I_i \) corresponds to smaller atoms \( g_{\gamma} \), \( \gamma \in I_i \), within a time support included in the support of \( g_i \), and modulated at the same frequency.

Let suppose that the atom \( g_i \) is chosen in a pursuit. \( Rf = f - C(f,g_i)p_i \) becomes the residue of this pursuit on the signal \( f \). For all \( \gamma \in I_i \), \( \langle Rf, g_{\gamma} \rangle \) represents the amount of "energy" of \( Rf \) located on the time-frequency support of \( g_{\gamma} \).

This amount must be smaller than the signal "energy" \( \langle f, g_{\gamma} \rangle \) at the same location. Moreover the corresponding decrease \( C(f,g_i)p_i \) of signal energy cannot be greater than the initial signal energy itself. This is formalized in Equations (6) and (7):

\[
|\langle Rf, g_{\gamma} \rangle| \leq |\langle f, g_{\gamma} \rangle|, \quad (6)
\]
\[
|\langle C(f,g_i)p_i, g_{\gamma} \rangle| \leq |\langle f, g_{\gamma} \rangle|, \quad (7)
\]

From these relations, we derive the new correlation function \( C(f,g_i) \), which maximizes the amount of signal energy that the pursuit can grasp, when choosing the atom \( g_i \):

\[
C(f,g_i) = \min_{\gamma \in I_i} \frac{|\langle f, g_{\gamma} \rangle|}{|\langle f, g_{\gamma} \rangle|} \quad (8)
\]

where \( \epsilon \) is evaluated as follows:

- if \( \langle f, g_{\gamma} \rangle > \) have the same sign, for all \( \gamma \in I_i \), then \( \epsilon \) is this common sign.
- else \( \epsilon = 0 \).

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 Figures 2-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 2-b, MP introduces a pre-echo effect by choosing atoms that overlap the attack time-location, whereas HRMP does not choose any such atom.

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 frequency-resolution is thus decreased but it performs a higher time-resolution decomposition than MP.

However for such audio applications as attack pattern recognition or precise tracking of partials,
the most important is to keep a good localization of the attacks, because the ear is very sensitive to transients: hearing the attack of a musical instrument is often almost sufficient to identify it.

6 Summary

HRMP provides a time-frequency representation adapted to the specificities of sound signals. Moreover, the signal representation HRMP provides is easily related to perceptual entities (transients, partials, clicks, ...). Thus, HRMP allows more precise or selective sound processing. We have presented, for example, its ability to process separately the sustained and transient parts of a piano sound. We are also considering the ability of the method to extract easily the parameters of formant waveforms synthesizers (central frequency, amplitude, bandwidth, and especially excitation duration).

References


