MULTI-RESOLUTION PARTIAL TRACKING WITH MODIFIED MATCHING PURSUIT

Pierre Leveau [1],[2]

[1] GET - ENST (Télécom Paris) - TSI
46, rue Barrault
75634 Paris Cedex 13 - France
pleveau@tsi.enst.fr

Laurent Daudet [2]

[2] Laboratoire d'Acoustique Musicale
11, rue de Lourmel
75015 Paris - France
daudet@lam.jussieu.fr

ABSTRACT

The widely used Matching Pursuit algorithm processes the input signal as a whole, and as such does not build relationships between atoms that are selected at every iteration. For audio signals, variants of this algorithm have been introduced that catch structured sets of atoms ("molecules") sharing common properties: harmonic relationship, time-frequency proximity. However, they are limited by the use of a single scale, hence a fixed time-frequency resolution, within a molecule.

In this study, we propose a modified Matching Pursuit that groups atoms at different scales within a given frequency line, allowing molecules with an optimized time and frequency resolution. Results on simple signals, as well as real audio recordings, show that the extra flexibility provided by multiresolution comes at a small computational cost.

1. INTRODUCTION

The standard Matching Pursuit algorithm (MP) [1] is a greedy algorithm giving an approximation of a signal x as a linear combination of M elementary waveforms ("atoms") $\mathbf{w}_m$ chosen within an overcomplete dictionary $\mathscr{D}$ of N atoms:

$$\mathbf{x} = \sum_{m=1}^{M} \alpha_m \mathbf{w}_m \qquad \text{where } \mathbf{w}_m \in \mathscr{D}. \tag{1}$$

The MP decomposition proceeds as follows. At every iteration, the atom the most correlated with the signal (*i.e.* with the largest scalar product $\alpha_i = |\langle \mathbf{w}_i, \mathbf{x} \rangle|$) is selected, and the corresponding weighted atom $\alpha_{i_0}\mathbf{w}_{i_0}$ subtracted from the signal. This process is iterated (which implies an update of all scalar products) until some stopping criterium is reached, *e.g.* on the Original-to-Residual energy Ratio (ORR). For music or speech signals, Gabor atoms of different scales are generally employed, and MP has been successfully applied to various signal processing problems such as audio coding or blind source separation.

In its standard form, MP usually gives atoms at uncorrelated time-frequency locations between consecutive iterations. This means that partials will not be extracted as a whole, leading to potentially poor reconstruction when only few atoms are employed : this leads to amplitude modulations. To overcome this issue, so-called molecular variants of the Matching Pursuit algorithm have been introduced [2, 3, 4]. They exploit natural organizations of the atoms in the audio signals, such as harmonic relationships, time-frequency or time-scale proximities. The partial tracking aspect has been handled in a mono-resolution context with local cosines [3] or Gabor atoms [4]. Partial tracking algorithms have also been developed in the past years with other approaches. Some of them are based on the Short Time Fourier Transform [5, 6]: these methods consequently have a fixed time-frequency resolution. Levine [7] uses an octave filter-bank to perform bandwise sinusoidal modeling, allowing a high time resolution in high frequencies, and a high frequency resolution in the low frequencies. In his study, the multi-resolution aspect has proven efficient in reducing pre-echo artefacts.

In this paper, a method that transposes the molecular concept into a *multi-resolution* context is proposed: the Multi-Resolution Molecular Matching Pursuit (MRMMP). Multi-resolution is a very useful feature as it allows a simultaneously sparse representation of steady-state and transient parts of the signals ; however if not carefully designed, this increase in dimensionality could lead to overly complex decomposition schemes. In section 2, the details of the algorithm are presented. Then, in section 3, we present and comment the results that emphasize the extra flexibility provided by our algorithm, on a single sinusoid and a real audio signal. Section 4 concludes with future improvements.

2. MULTI-RESOLUTION MOLECULAR MATCHING PURSUIT

Basically, the only implementation difference between standard MP and Multi-Resolution Molecular Matching Pursuit (MRMMP) resides in the set of allowed indexes for the search of the atom maximally correlated with the signal. In MP, the search is always performed on the whole dictionary. In MRMMP, two types of searches are performed : in the case of *seed* atoms, the search is global as in MP; otherwise, the search is restricted to a local search, in order to find atoms that are maximally close in the time-frequency domain.

In this study, the dictionary $\mathscr{D}$ is composed of Gabor atoms $\mathbf{w}_{(s,u,\omega)}$ with 3 parameters corresponding to scale s, time u and frequency ω :

$$\mathbf{w}_{(s,u,\omega)}(t) = \frac{1}{\sqrt{s}} w\left(\frac{t-u}{s}\right) e^{2i\pi\omega t} \tag{2}$$

This work was supported by the French Ministry of Research and Technology, under contract ACI "Jeunes Chercheuses et Jeunes Chercheurs" number JC 9034.

2.1. Seed atom selection

The first atom is searched in the same way as in the standard MP algorithm. It is the *seed* atom of the first molecule, with parameters s_e,u_e,ω_e.

2.2. Consecutive atom selection

Following the selection of a seed atom (s_e,u_e,ω_e), the candidate atoms at the next iteration are restricted to a small subset $\mathscr{I}$ of the whole dictionary $\mathscr{D}$, with locality constraints as follows. Only candidates with the same frequency ω_e are considered (search restricted to pure tones). N_s atoms, where N_s is the number of scales inferior or equal to current scale s_e , are considered forward in time, one at each scale. At a given scale s_i, the forward candidate atom is located at time $u_i^f = u_s + \frac{s_i}{2}$. Similarly, N_s atoms are considered backwards in time, with, at scale s_i, a time location $u_i^b = u_s - \frac{s_i}{2}$.

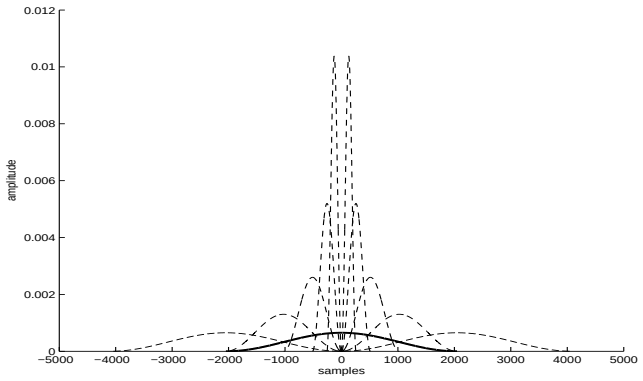


Fig. 1. Candidates for backward and forward extensions of a molecule (here for the second atom of the molecule). Thick line: envelope of the molecule seed ($s_e = 4096$ samples), dashed lines: envelopes of the candidates ($s_i = 256, 512, 1024, 2048, 4096$).

This choice of distance between two consecutive atoms avoids hole forming in the partial, that is observed in standard MP (see Section 1). Indeed, the 50 % overlap between neighboring atoms at the same scale guarantees perfect reconstruction of constant-amplitude sines using Gabor atoms with a Hanning window. When the best matching atom is selected within this subset $\mathscr{I}$, the corresponding waveform $\alpha_{i_0}\mathbf{w}_{i_0}$ is subtracted from the signal as in MP.

Following iterations are done in the same manner except that the backward candidate locations are computed as a function of the location of the first atom of the molecule in progress, and the forward candidate as a function of its last atom. The molecule is terminated when the selected atom reaches a fixed proportion of the first atom or an absolute energy threshold. The choice of this criterion is arbitrary. After each terminated molecule, the algorithm looks for the next seed atom until the general stopping criteria has been reached.

Note that with this method, groups of adjacent time-frequency atoms representing an given partial are considered as a whole, which is a definite improvement for modelling the partials in the signal.

2.3. Optimization of the atom weights

As an additional benefit, this atom grouping allows for some (optional) optimization. Once all atoms within a molecule have been selected, their weights α_i can be refined as follows. Given a set S_m of i_m atoms $\mathbf{w}_i$ that form the molecule m, find

$$\{\alpha_i^{opt}\}_{i=1..i_m} = \arg\min_{\alpha_i, i=1..i_m} ||\mathbf{x} - \sum_{i=1}^{i_m} \alpha_i \mathbf{w}_i||$$

With $\mathbf{W}$ the matrix of the molecule's atoms in column, the least-square solution $\{\alpha_i^{opt}\}_{i=1..i_n}$ is given by be the matrix of the atoms in column, α^{opt} the column vector containing the optimal weights . Then:

$$\alpha^{opt} = (\mathbf{W}^H\mathbf{W})^{-1}\mathbf{W}^H\mathbf{x}$$

After optimizing these weights, an update of scalar products has to be performed.

Note that in our case the column vectors (or the atoms) are not strongly correlated, since their temporal support are distinct between each other. It implies that the matrix $\mathbf{W}^H\mathbf{W}$ is well-conditioned and that numeric computations lead to robust results. This optional step is related to the so-called orthogonal MP [8], where at each iteration the signal is optimally projected on the span of *all* previously selected atoms; in MRMMP this optimization is *local* to within a molecule, *i.e.* to where atoms are most correlated, with significant computational savings.

2.4. Computational load

The computational cost, usually dominated by the update of the scalar products, is reduced when compared with MP. Within a molecule, these update steps are quick: only the scalar products related to candidate atoms need an update (N_s updates). This is done by directly computing the inner product between the atom and the signal. Given the candidate scale s_i, this lead to a computational load of $O(\sum s_i)$ ($\sum s_i$ multiplications and $\sum s_i$ additions for each candidate), instead of $O(\sum(s_i \log(s_i)))$ when a Fast Fourier Transform algorithm is employed. When a molecule is terminated, all the inner products touched by the molecule time support must be updated ; here only this step has similar complexity as standard MP.

3. EXPERIMENTS

3.1. Single sinusoid

The described algorithm has been applied on a single sinusoid at frequency 1000 Hz (sample rate 44.1 kHz), lasting about 1s. The scales are 256, 512, 1024, 2048, 4096 samples ($N_s = 5$) and the window shift 128 samples. The FFT have been computed on zero-padded windows in order to have the same frequency sampling for each block. The stop criterion has been fixed at 24 atoms, which is the size of the first molecule with the MRMMP algorithm. Three algorithm have been tested: the MP, the MRMMP without molecule optimisation, the MRMMP with molecule optimisation. On Figure 2, the reconstruction and the residual resulting from the 3 methods are displayed. The respective ORRs are 34.6, 36.2 and 50.0 dB.

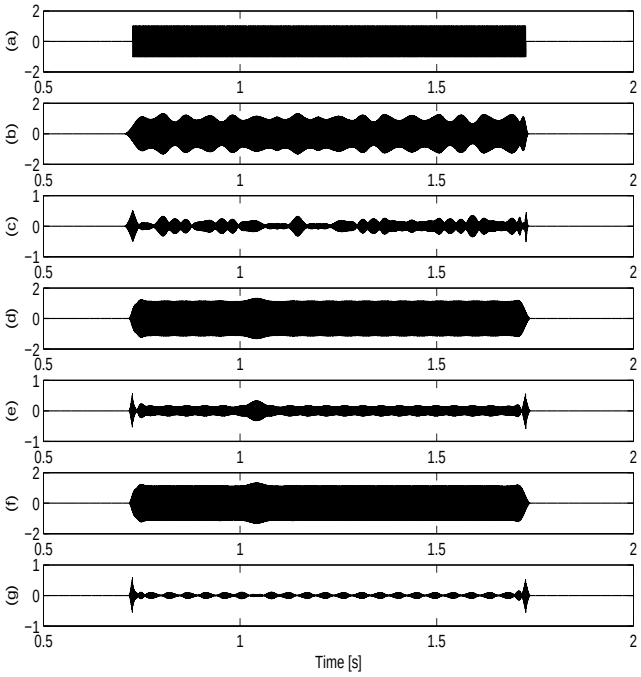


Fig. 2. Comparison between the MP and MRMMP algorithms. (a): original signal. (b), (d), (f): constant amplitude sinusoid reconstructed respectively with MP atoms, MRMMP atoms without optimization, MRMMP with optimization. (c), (e), (f): respective residuals.

The standard Matching Pursuit (Figure 2 (b) and (c)) algorithm fails to efficiently represent the partial: the first selected atoms have a maximum amplitude that exceeds the amplitude of the sinusoid. Here, MP adds energy in the middle of the atom (bumps in (b), sounding as amplitude modulations).

The MRMMP without optimisation shows a more advantageous behaviour: the first atom is (as in MP) too energetic (see the bump on the reconstructed signal (d)). Then the next atom takes less energy, and the following ones converge to a steady value. However this value does not lead to a perfect reconstruction in the steady part of the sinusoid. In the beginning and end of the signal we see the advantage of using a multiresolution approach : a too long atom extracted near the beginning of the partial is compensated by the extraction of a smaller scale atom with an opposite phase, hence a reduction of pre-echo. A very interesting remark arises when looking at the new ORR: 36.2 dB $>$ 34.6 dB. This is apparently paradoxical: each iteration of MRMMP is $\mathscr{L}_2-$suboptimal compared to MP (since it searches the best matching atom in a subset of $\mathscr{D}$), so one would naively expect always better results for MP. However, after a number M of iterations, MRMMP leads (in this particular case) to better results ! Remember that MP is only optimal *at every iteration*, and generally there is no guarantee that after M iterations it gives the best M-atoms approximation. In turns out that the guided search in MRMMP can indeed lead to better global results.

Finally, the least square optimisation of the molecule is efficient in reducing the residual in the steady part of the sinusoid. However, artifacts are unavoidable at the end and at the beginning. Indeed, at the sides, the scale changes between consecutive atoms; since windows at different scales are homothetic (Hanning windows in our case), the overlap cannot lead to perfect reconstruction.

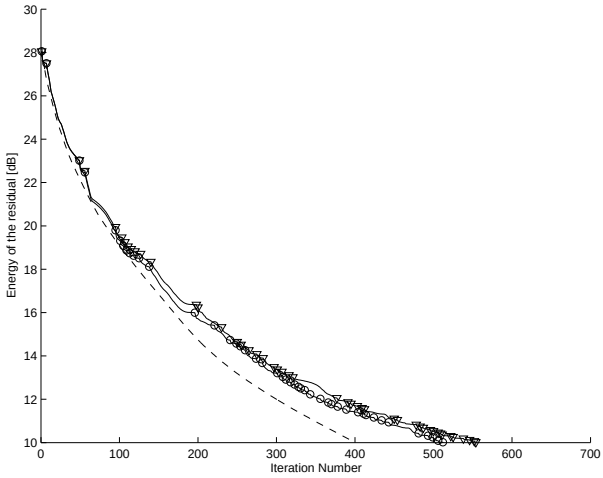


Fig. 3. Energy of the residuals as a function of the number of iterations. Solid line with triangles: MRMMP without optimisation, with circles: MRMMP with optimisation (triangles and circles mark the molecule starts). Dashed line: MP.

3.2. Real signals

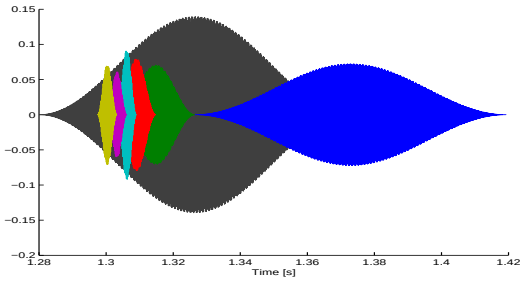


Fig. 4. Example of a glockenspiel partial modeled by 7 atoms of different scales, obtained with MRMMP.

The performances in terms of ORR of the MRMMP on a glockenspiel recording is shown on Figure 3. The stop criteria for the molecules have been set to 20dB below the first atom of the molecule, and 40 dB below the first atom of the whole MRMMP algorithm. The glockenspiel signal contains well defined partials with constant frequencies. An example of a partial modeled by several atoms is shown on Figure 4. The MP decay curve is displayed for reference purpose: as stated above the goal of MRMMP is not to achieve the best ORR but it remains close to it. However the main improvement when compared to MP is to catch perceptually relevant structures. Note that the local optimization significantly improves the overall ORR for a given number of iterations (up

to 1 dB).

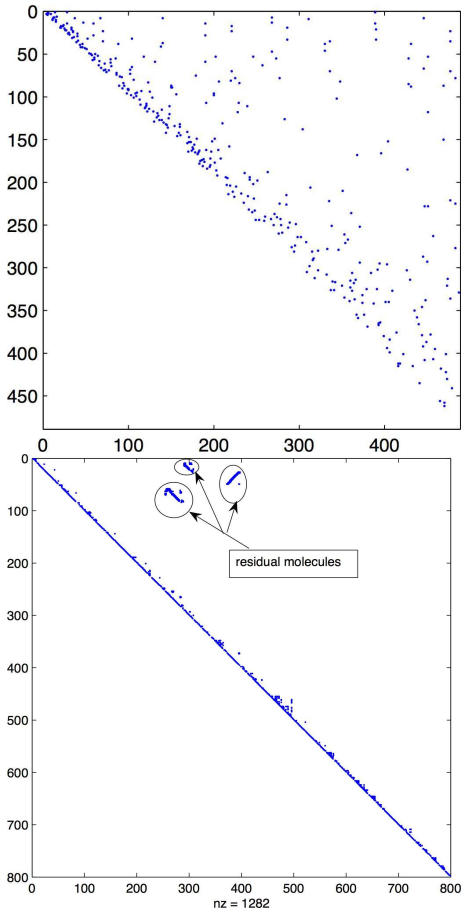


Fig. 5. Adjacency matrix for the MP (top) and the MRMMP (bottom) algorithms. Residual molecules are circled.

Figure 5 shows the adjacency matrix of the SMP and the MRMMP atomic representation of the glockenspiel signal: if atom i and atom j have overlapping temporal support and the same frequency, they are said to be adjacent and a dot is displayed at coordinates (i, j) .

Unsurprisingly, MP shows few interconnections between consecutive atoms when compared to the MRMMP, which is designed to extract adjacent atoms. However, an interesting feature of the MRMMP also appears on this figure: the molecules resulting from the residual of another molecule can be easily identified because each of its atoms is adjacent to one of the original molecule, and atoms are mutually adjacent inside the residual molecule. If the goal of the sparse approximation is audio coding or partial extraction for analysis-synthesis, these low energy artefacts from significant molecules can be removed because of their low perceptual significance. They can otherwise be allocated to the original molecule if a more precise description of the partial is needed.

4. CONCLUSION

In this paper, a new method is proposed for high-resolution partial tracking. It is a modification of the Matching Pursuit algorithm, adapted to audio signals in the way that it catches some of its relevant structures. The novelty of this paper is the inclusion of multiresolution, an important feature for the extraction of amplitude-modulated sinusoids and transient parts of the signals. We have shown it is possible to design practical algorithms where that the extra flexibility provided by multiresolution comes at a small computational cost. In some simple cases, MRMMP, although suboptimal at each iteration as compared to MP, can even lead to better ORR results. Finally, the resultant structured representations can be a relevant starting point for higher-level processing, such as sound transformation and automatic transcription.

Towards a more complete bottom-up processing of audio-signals, the presented approach will be extended to the extraction of more complex but relevant structures. Molecules of harmonic atoms [2] can be built in the same manner. Chirped Gabor atoms could also benefit from a local candidate atom selection, by examining frequency proximity in addition to time proximity. From a practical point of view this method will be implemented in MPTK, a fast open-source Matching Pursuit toolkit [9].

5. ACKNOWLEDGMENTS

The authors would like to thank R. Gribonval, S. Krstulović and G. Richard for useful discussions.

6. REFERENCES

[1] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionnaries," *IEEE Trans. on Signal Proc.*, vol. 41, pp. 3397–3415, 1993.

[2] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Trans. Signal Proc.*, vol. 51, no. 1, pp. 101–111, jan 2003.

[3] L. Daudet, "Sparse and structured decompositions of signals with the molecular matching pursuit," *IEEE Trans on Audio, Speech and Language Proc.*, to be published in 2006.

[4] S. Krstulovic, R. Gribonval, P. Leveau, and L. Daudet, "A comparaison of two extensions of the matching pursuit algorithm for the harmonic decomposition of sounds," in *Proc. WASPAA*, 2005.

[5] M. Lagrange, S. Marchand, and J.-B. Rault, "Tracking partials for the sinusoidal modeling of polyphonic sounds," in *Proc. ICASSP*, 2005.

[6] P. Depalle, G. Garcia, and X. Rodet, "Tracking of partials for additive sound synthesis using hidden markov models," in *Proc. ICASSP*, 1993.

[7] S. N. Levine, T. S. Verma, and J. O. Smith III, "Multiresolution sinusoidal modeling for wideband audio with modifications," in *Proc. ICASSP*, 1998.

[8] Y.C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. on Signals, Systems and Computers*, 1993.

[9] S. Krstulovic and R. Gribonval, "The matching pursuit toolkit (mptk), http://gforge.inria.fr/projects/mptk/," .