MODEL-BASED MATCHING PURSUIT - ESTIMATION OF CHIRP FACTORS AND SCALE OF GABOR ATOMS WITH ITERATIVE EXTENSION

*Pierre Leveau, Laurent Daudet**
Laboratoire d'Acoustique Musicale
Univ. P&M Curie/Paris 6 – 11 r. Lourmel – 75015 Paris, France
{leveau,daudet}@lam.jussieu.fr

*Sacha Krstulović, Rémi Gribonval**
IRISA
Campus de Beaulieu – 35042 Rennes Cedex, France
{sacha,remi}@irisa.fr

ABSTRACT

By definition, the Matching Pursuit algorithm with constant (or "flat") Gabor atoms provides a coarse estimate of frequency modulated sinusoids in music and voice signals. Chirped Gabor atoms, closer to the nature of these signals, would fit them in a finer and sparser way. Though a method for the direct analytic estimation of chirped Gabor atoms has been proposed in the past [1], the present article proposes an alternative method where the chirp factor and scale parameter are estimated through a regression over an iteratively selected chain of small-scale atoms defined by a Short Time Fourier Transform. This new technique suits the Matching Pursuit framework, and is compared with a "flat atoms" version of the algorithm. The influence of various frequency interpolation techniques over the sparsity of the resulting representation is also studied.

1. INTRODUCTION

The Matching Pursuit (MP) algorithm decomposes signals into a sum of atoms. These atoms are selected iteratively: at each iteration, the atom maximizing the correlation with the signal is extracted. The resulting approximation of the signal can be written:

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

In the case of frequency modulation, an atom can be written:

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

where s is the scale, u the onset time of the atom, ω the frequency and c the chirp rate.

As shown in [1], the extensive search of the best chirped Gabor atom is very costly. Thus an optimisation technique has been proposed to find the *locally best* chirped atom. First the best flat atom is searched, then the chirped factor and scales are reestimated using the local behaviour of the signal.

Our study presents a variant of the method evoked above, using a local optimisation of all four atom parameters from a flat gabor atom. The technique involves the extension of a small scale flat Gabor atom. It is inspired from the classical partial tracking algorithms based on the Quadratic Interpolated Fast Fourier Transform (QIFFT) [2], and uses a linear regression to estimate chirp factor and frequency parameters.

The estimation of the atom parameters is first detailed, then we include it in a MP algorithm and compare it to MP with only flat atoms.

2. ESTIMATION OF THE ATOM PARAMETERS

Each iteration of the MP algorithm begins with the search for the small-scale flat atom which is the most correlated with the signal. Concretely, it is equivalent to search for a maximum in a Short Time Fourier Transform (STFT) modulus time-frequency representation of x, stored in an array G (namely a *Gabor block*) . Given (u, ω) the time and frequency of the maximum, we extend iteratively the atom while the correlation between the atom and the signal increases. The extension step is Δs, equal to the hop time of the STFT. The extension is done successively once forward, once backward.

2.1. Algorithm

1. **initialisation**: f is initialized to f_0, c to 0 and s to the Gabor atom scale of the STFT representation s_w.
2. do:

*This joint work was partially supported by the MathSTIC program of the French CNRS.

Forward extension: Starting from a (u_0, ω_0) time-frequency point, the maximum of the Gabor block at time $u_0 + \Delta s$ within the following authorised frequency range is searched:

$$f_1 \in [f_0 - f_0.C.\Delta s, f_0 + f_0.C.\Delta s] \quad (3)$$

where C is the maximum allowed chirp rate (1 octave/s is convenient for music signals). In the numeric application, this range is represented by a vector of discrete frequencies. The borders of this vector are the one of the range defined in 3 rounded to the closest sampled frequencies outside this range.

To have a reliable estimation of f_1, two solutions can be imagined: to use a high zero-padding factor (equivalent to an interpolation with a sinus cardinal), or a quadratic interpolation of the frequency peak (as used in QIFFT). The efficiency of this techniques will be discussed in the next section. If the estimated frequency is outside the authorised frequency range, f_1 is set to its closest border (see Figure 1).

We now have a time vector $\mathbf{T}$ and a frequency vector $\mathbf{F}$ containing the coordinates of the consecutive maxima of the STFT. The new chirp and frequency parameters $\hat{c}$ and $\hat{f}$ are estimated using a linear regression. Indeed, the following hypothesis is made:

$$\mathbf{F} = c(\mathbf{T} - T_1) + f + \epsilon$$

where T_1 is the first element of vector $\mathbf{T}$ and ϵ the error.

$\hat{s}$ is obtained by incrementing s by Δs, $\hat{u}$ is equal to u: the onset time of the atom is not modified in the forward case. Then, the correlation product is computed:

$$P = | < x_{seg}, g_{\hat{u},\hat{f},\hat{s},\hat{c}} > |^2$$

where x_{seg} is the truncature of x on the atom location.

It is compared with the scalar product of the Gabor atom generated with the previous estimations. If the new parameters lead to a more correlated atom, they are kept. Otherwise, the previous parameters are restored and the extension has failed.

- **Backward extension** The same procedure is performed backward. The only difference is in the onset time update: $\hat{u}$ is set at $u - \Delta s$.

3. **Termination:**The extensions are stopped if a forward then a backward extension have failed, or when a maximum scale has been reached (e.g. 0.1 s). Indeed allowing very long atoms can lead to higher scalar products, but it can bring musical noise to the resynthesised signal.

An example of the whole process is shown on Figure 1, where 34 extensions have been performed to get the final atom.

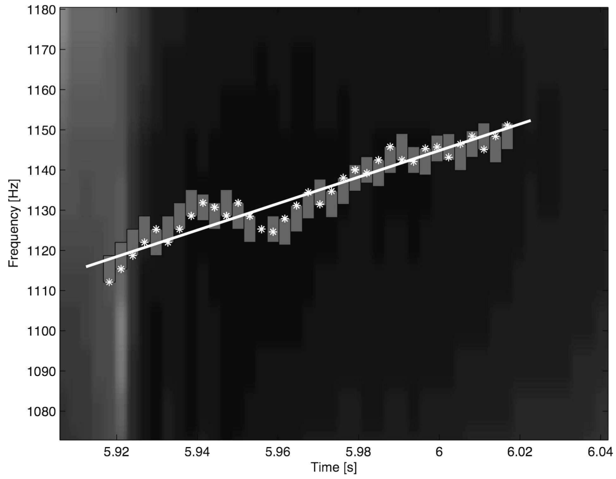

Fig. 1. Extension of a single Gabor atom on a trumpet partial: in background, small window STFT representation; successive frequency search domains (gray rectangles); peak frequencies (white stars); estimated chirp (white line).

2.2. Computational cost

When compared with an exhaustive search of the chirped Gabor atoms, the computational asset of this method resides in two aspects. First, as in [1], the chirp parameter is estimated and not selected among a huge dictionary of pre-calculated atoms. Second, the scale parameter is also estimated allowing to avoid the heavy computational load of initialising and updating Gabor blocks with different window sizes, especially if they are computed for each of the reachable scale by our algorithm (e.g. from 512 samples to 8192 samples with a step of 128 samples).

3. EXPERIMENTS

3.1. Influence of zero-padding

The parameter estimation has been implemented in a modified MP algorithm. It has been applied on two audio signals: a trumpet sample and a singing voice sample, each lasting 6 seconds. These sound files contain frequency modulations, and thus should benefit from the chirp factor estimation. In the following experiments, the window size s_w is 512 samples, and the hop time Δs is 128 samples. The stop criterion is when Original-to-Residual Ratio (ORR) reaches 20 dB, and

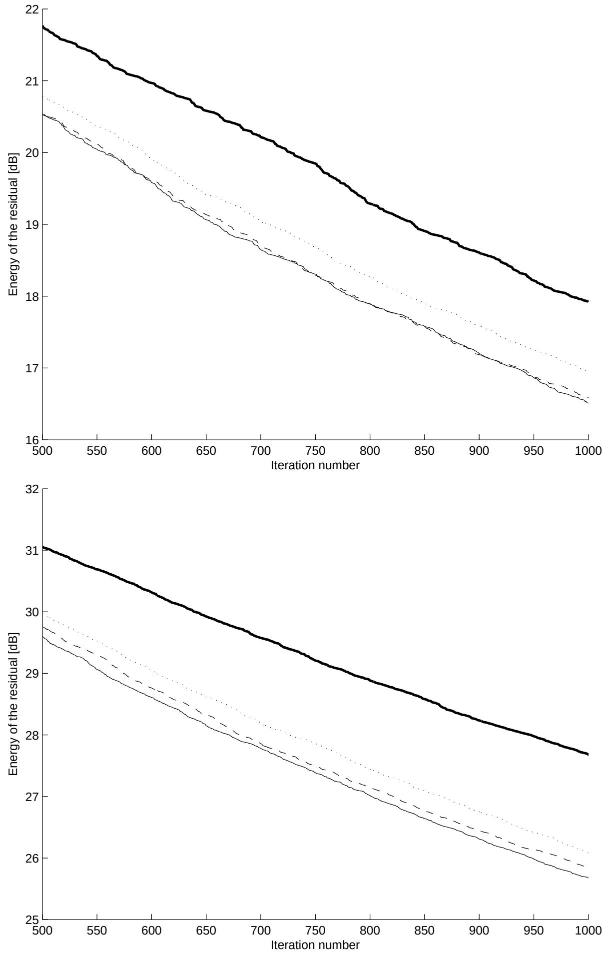

Fig. 2. Influence of zero-padding: ORR curves for iterations 500 to 1000 of the MP algorithm showing the (top: trumpet sample, bottom: voice sample). Thick line: chirp factor constrained to 0, thin lines: estimated chirp factor (dotted: $N = 2048$, dashed: $N = 4096$, solid: $N = 8192$).

no quadratic frequency interpolation has been performed. The influence of the FFT size has first been studied. On Figure 2, the energy of the residual on the trumpet and the voice are displayed for 3 FFT sizes (2048, 4096, 8192), as a function of the iteration number.

A version of our algorithm without allowed frequency modulation has also been tested: in this case, the extension is only performed within a single frequency bin, and the chirp factor c is constrained to 0. However, this algorithm is not equivalent to a generic Matching Pursuit algorithm with multi-resolution flat atoms, where the scale of the atom is not estimated but comes at the selection step. In our case, at each iteration, the originally selected flat atom has a fixed small scale, then is extended.

It must be pointed out that the ORR curves of the experiments with c constrained to 0 are represented by only one curve on Figure 2: in this case the MP algorithm leads to very close results for any zero-padding factor. Indeed, the difference between the three ORRs is at most 0.05 dB at any iteration.

On the contrary, when the chirp factor is estimated, the zero-padding factor has a stronger influence: higher FFT sizes lead to lower ORRs for a given iteration number. This shows the additional information brought by the additional c parameter. For the late iterations, the difference between $N_{FFT} = 2048$ and $N_{FFT} = 8096$ is 1dB for the singing voice, 0.2 dB for the trumpet.

3.2. Influence of quadratic frequency interpolation

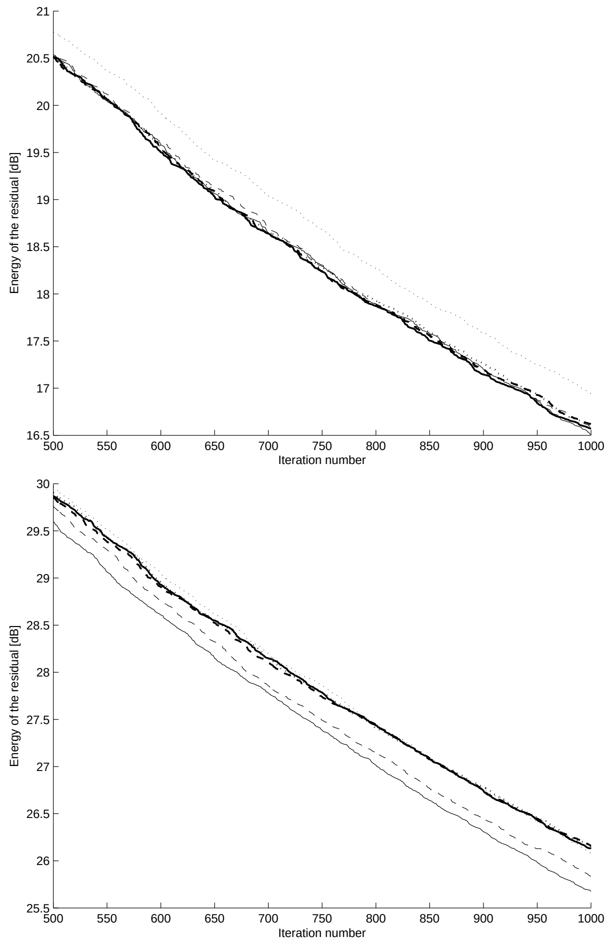

Fig. 3. Influence of the quadratic frequency interpolation: ORR curves for iterations 500 to 1000 of the MP algorithm (top: trumpet sample, bottom: voice sample). Thick line: with quadratic frequency interpolation, thin lines: without quadratic frequency interpolation. dotted: $N = 2048$, dashed: $N = 4096$, solid: $N = 8192$.

The influence of quadratic frequency interpolation on the parameter estimation has been studied in the same MP framework. Again, the three FFT sizes are 2048, 4096 and 8192. The ORR curves are shown on Figure 3 for the same sound samples.

The quadratic interpolation leads to close results for the 3 FFT sizes: for a given iteration, the differences between the ORRs do not exceed 0.1 dB. However, it does not necessarily lead to better results than the parameter estimation without quadratic frequency interpolation.

For the trumpet sample, the improvement is significant for $N_{FFT} = 2048$, achieving performances close to the $N_{FFT} = 8192$ case without quadratic interpolation. For the singing voice sample, the quadratic interpolation only improves the ORR for the first 30 iterations, then it reduces the performances, slightly for $N_{FFT} = 2048$ but drastically for $N_{FFT} = 8096$.

These two examples show that the quadratic interpolation does not improve the ORR in every case, but reduces the influence of the choice of the zero-padding factor. Hence, quadratic interpolation can be used with a small zero-padding factor if a small computational cost is needed. If the objective is a low ORR, to set a high zero-padding factor without quadratic frequency interpolation is the better solution.

4. CONCLUSION

In this study we introduce a new method to extract chirped Gabor atoms. It involves a local optimisation of the atom parameters by extending small-scale flat atoms. This method catches rapidly perceptually relevant structures. Its computational asset is that only one small-scale Gabor block is used to estimate multi-scale atoms. The update of the scalar product and the search for the maximum correlation is thus much lighter. This technique can also be seen as a reduction of the number of parameters used to describe a frequency modulation: instead of using a possibly large number of small scale atoms to represent this phenomenon, a single large scale chirp atom is estimated on the basis of such small atoms.

The following work will deal with the extraction of *harmonic chirps* in the Matching Pursuit framework, to go further in the search for perceptually relevant features in the music signal. The chaining of this type of atoms will also be investigated to catch longer structures (*molecules*) like partials, as a continuation of [3] and [4]. The chirplet chain approach presented in [5] will also be envisaged.

5. REFERENCES

[1] R. Gribonval, “Fast matching pursuit with a multiscale dictionary of gaussian chirps,” *IEEE Transactions on Signal Processing*, 2001.

[2] X. Serra, *A System for Sound Analysis/Transformation/Synthesis based on a Deterministic plus Stochastic Decomposition*, Ph.D. thesis, Stanford University, 1989.

[3] 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.

[4] P. Leveau and L. Daudet, “Multi-resolution partial tracking with modified matching pursuit,” in *Proc. ICASSP*, 2006, submitted.

[5] E. Chassande-Mottin, “Meilleures chaînes de chirplets et ondes gravitationnelles,” in *Journées ”Décompositions adaptatives” du GdR–CNRS ISIS (Marseille, France)*, 2004.