A Class of Algorithms for Time-Frequency Multiplier Estimation

Anaïk Olivero, Bruno Torrésani, Richard Kronland-Martinet

A Class of Algorithms for Time-Frequency Multiplier Estimation

Anaïk Olivero, Bruno Torrésani, *Senior Member, IEEE,* and Richard Kronland-Martinet, *Senior Member, IEEE,*

***Abstract*—We propose here a new approach together with a corresponding class of algorithms for offline estimation of linear operators mapping input to output signals. The operators are modelled as multipliers, i.e. linear and diagonal operator in a frame or Bessel representation of signals (like Gabor, wavelets ...) and characterized by a transfer function. The estimation problem is formulated as a regularized inverse problem, and solved using iterative algorithms, based on gradient descent schemes. Various estimation problems, which differ by a choice for the regularization function, are studied in the case of Gabor multipliers. The transfer function actually provides a meaningful interpretation of the differences between the two signals or signal classes under consideration, and examples are discussed. Furthermore, examples of signal transformations with such Gabor transfer functions are also given.**

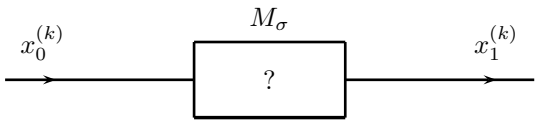
***Index Terms*—Analysis/Transformation/Synthesis, Frame Representations, Frame Multipliers**

I. Introduction

Analysis/Transformation/Synthesis is a general paradigm in signal processing, that aims at manipulating or generating signals for practical applications like signals transformation, compression, denoising or source separation. Analysis/Transformation/Synthesis is often performed starting from a parametric signal model (for example the sinusoids+noise model [1], [2]). The parameters are first estimated, and a new signal is synthesized from the modified parameters. Analysis/Transformation/Synthesis can also be performed starting from a linear signal representation [3] (for example a short time Fourier representation, or a Gabor or wavelet expansion), by directly modifying the coefficients prior to resynthesis.

In this context, a signal transformation can be constructed by pointwise multiplication of the analysis coefficients of the signal representation and a *transfer function*. Such transformations are generically called *multipliers*. Denoting by $\boldsymbol{\sigma}$ the transfer function we shall denote by $\mathbb{M}_{\boldsymbol{\sigma}}$ the corresponding multiplier. In case of the Fourier representation, this obviously corresponds to a time-invariant linear filter, i.e. a convolution operator. Other signal representations defined in the general context of frames or Bessel sequences also lead to frame or Bessel multipliers. When time-frequency representations are used, such operators give the possibility of implementing time-varying linear filters [4], [5]. Applications of such multipliers can also be source separation [6], where binary Gabor multipliers are used to select a subset of time-frequency indices in the Gabor domain. Here, we will exploit such Gabor multipliers in the context of sound transformations.

The first goal of this paper is to study a general method to estimate a transfer function between two signals x_0 and x_1 or families of input-output pairs $(x_0^{(k)}, x_1^{(k)})$, given a signal representation. In other words, given x_0 and x_1 (resp. the pairs $(x_0^{(k)}, x_1^{(k)})$) and a signal representation, how to estimate the transfer function $\boldsymbol{\sigma}$ such that $\mathbb{M}_{\boldsymbol{\sigma}} x_0$ is closest to x_1 (resp. $\mathbb{M}_{\boldsymbol{\sigma}} x_0^{(k)}$ is closest to $x_1^{(k)}$ on average) ?



This estimation problem has been first addressed in [7] in an approximated formulation in the case of Gabor frames. Then, we have proposed in [8] the use of an iterative algorithm to estimate a Gabor multiplier. The main contribution of the present paper is to extend the class of algorithms to estimate multiplier in the more general framework of frames, and give interpretations and audio examples to emphasize the relevance of our approach. With suitable choices of the regularization, the problem is convex and we use (provably convergent) iterative strategies based on gradient schemes for its numerical resolution. We also consider some other choices of regularization of practical interest, which yield non-convex problems. We also show that our formulation allows considering more complex models for signal transformations, such as multiple multipliers, i.e. linear combinations of multipliers. The problem we address can be cast as an offline system identification problem, solved in the spirit of [9], [10]. However, we rather focus in this paper on applications to sound analysis, categorization and synthesis.

A first application of our approach is in the domain of sound analysis. Sounds categorization can be performed on the basis of pairwise comparisons. Indeed, an estimated transfer function can be viewed as a descriptor of the differences between two signals [11], [7] or two signal classes under consideration (instrument classes in [12], [13]). We have shown in [12] that such descriptors could yield sensible classifications within controlled musical signal families. A second important application of the approach is directly related to the possibility of synthesizing new sounds given an input sound and an output

Manuscript received September 5, 2011.

Anaik Olivero and Bruno Torrésani are with Laboratoire d'Analyse, Topologie et Probabilités, Aix-Marseille Université, 39 rue F. Joliot-Curie, 13453 Marseille cedex 13, France. (e-mail: olivero@cmi.univ-mrs.fr, torresan@cmi.univ-mrs.fr).

R. Kronland-Martinet is with the NCRS-Laboratoire de Mécanique et d'Acoustique (LMA), 13402 Marseille Cedex 20, France (e-mail: kronland@lma.cnrs-mrs.fr).

sound. For example, given an estimated mask $\boldsymbol{\sigma}$, consider a one-parameter family of masks $\boldsymbol{\sigma}_\tau$, $\tau \in [0,1]$, such that $\boldsymbol{\sigma}_0 = 1$ and $\boldsymbol{\sigma}_1 = \boldsymbol{\sigma}$. Acting on the input signal with multipliers with transfer functions $\boldsymbol{\sigma}_\tau$ yields a one-parameter family of signals that interpolate between input and output signals, i.e. a signal morphing. Developing such morphing schemes is one of the goals of our program [8], [14].

The paper is organized as follows. Section II is devoted to the general setting in which this work is done, and describes the basic concepts of signal representation and time-frequency analysis we shall be working with. Section III describes the proposed algorithms for multiplier estimation from a (family of) pair(s) of signals. Numerical results are presented and discussed in section IV.

II. Signals Representations

As mentioned above, a multiplier is defined by pointwise multiplication with a transfer function in a given representation space. For example, standard linear time-invariant filters are multipliers associated with the Fourier representation. Let us start by specifying the signal representations we shall be using. For the sake of simplicity, we limit the present discussion to 1D signals, considered as elements in (Hilbert) signal spaces, generically denoted by $\mathcal{H}$. Examples of interest are $\mathcal{H} = \mathbb{C}^L$ as well as the infinite-dimensional models such as $\mathcal{H} = \ell^2(\mathbb{Z})$. We are interested in representations in dictionaries, i.e. complete, parametrized sets of *atoms* $g_\lambda \in \mathcal{H}$. Given such a dictionary, under some additional conditions (see below), any signal can be characterized by the family of its *analysis coefficients* $\langle x, g_\lambda \rangle$. When the dictionary is an orthonormal basis of $\mathcal{H}$, the signal x expresses as

$$x = \sum_\lambda \langle x, g_\lambda \rangle g_\lambda \ . \tag{1}$$

In a more general framework, a signal representation can be constructed with an overcomplete family of atoms. With such families, signal representations such as (1) are not unique and several different sets of *synthesis coefficients* can generate a same signal x. For the sake of completeness, we introduce below the notions of frame and Bessel sequence before turning to the multiplier estimation problem. The choice of the decomposition will depend on the applications. In the context of audio signals, Gabor frames (subsampled version of Short Time Fourier Systems) are probably the most famous example of frame and will be studied in details in section IV. In image processing, wavelet basis [3] are largely used because they give a sparser representation of the signal. In audio applications, MDCT [15] is often used for audio coding, whereas Gabor frames are generally preferred for audio analysis applications.

A. Definitions

We first recall [16] some definitions, and conditions which leading to generative sequences of a separable Hilbert space $\mathcal{H}$.

- Let us denote by $G_\Lambda = \{g_\lambda : \lambda \in \Lambda\}$ a sequence of signals (or atoms) labelled by the index set Λ. G_Λ is a Bessel sequence if and only if there exists $0 < B < \infty$ such that

  $$\sum_\lambda |\langle x, g_\lambda \rangle|^2 \leq B \|x\|_2^2 \ , \ \forall x \in \mathcal{H}$$

  For such a Bessel sequence the analysis operator, denoted by $\mathcal{V}_g : \mathcal{H} \to \ell^2(\Lambda)$ is well defined, and reads

  $$\mathcal{V}_g x[\lambda] = \langle x, g_\lambda \rangle = X[\lambda]$$

  The synthesis operator is the adjoint of the analysis operator, denoted by $\mathcal{V}_g^\star : \ell^2(\Lambda) \to \mathcal{H}$, and defined as

  $$\mathcal{V}_g^* \alpha[l] = \sum_\lambda \alpha_\lambda g_\lambda[l]$$

  The frame operator is then given by $S = \mathcal{V}_g^* \circ \mathcal{V}_g$
- G_Λ is a frame if and only if there exist constants $0 < A \leq B < \infty$ such that

  $$A\|x\|_2^2 \leq \sum_\lambda |\langle x, g_\lambda \rangle|^2 \leq B\|x\|_2^2 \ , \ \forall x \in \mathcal{H} \ .$$

  This is equivalent to say that the frame operator is bounded below by A and above by B. A and B are respectively called the lower and the upper frame bounds. A perfect reconstruction of the signal x from the analysis coefficients is achieved because the frame operator is invertible, and allows the construction of the canonical dual frame sequence $H_\Lambda = \{h_\lambda = S^{-1} g_\lambda : \lambda \in \Lambda\}$, which leads to

  $$x = \sum_\lambda \langle x, g_\lambda \rangle h_\lambda = \mathcal{V}_h^* \circ \mathcal{V}_g \tag{2}$$

  The lower and upper frame bounds of the canonical dual frames are respectively given by $1/B$ and $1/A$.

These operators allow us to switch between the time representation of the signal and his transform domain representation. Note that except in the basis case, this correspondence is not one to one, and the range of the analysis operator is a proper closed subspace of $\ell^2(\Lambda)$.

B. Some signals representations

In the finite-dimensional situation $\mathbb{C}^L$, besides the time representation, the most familiar representation is provided by the Fourier basis, the analysis operator being the Discrete Fourier Transform (DFT). An example of frame is given by oversampled Fourier representations [17] that can be constructed by evaluating a DFT at $M > L$ frequencies. Let $m \in \{0, .., M\}$, an example of redundant Fourier representation of a signal x is given by

$$DFT_x[m] = \sum_{l=0}^{L-1} x[l] e^{-2i\pi bml/L} \ ,$$

where b is an integer. This construction of oversampled DFT implies that the linear independence between the Fourier atoms is lost.

Other standard representations which constitute a frame of signals are the time-frequency representations. The simplest example is provided by the short time Fourier transform,

which can be seen as the analysis map of a Gabor frame representation of the signal, as explained in [18]. A Gabor frame is an overcomplete family of time-frequency atoms g_{mn} generated by translation and modulation on a discrete lattice of a mother window $g \in \mathbb{C}^L$. These atoms can be defined as follows:

$$g_{mn}[l] = e^{2i\pi mb(l-na)/L} g[l-na] \ , \tag{3}$$

where a and b are two positive integers, such that L is multiple of both a and b. Here, all operations have to be understood modulo L. We set $M = L/b$ and $N = L/a$.

Extensions of Gabor frames are the non-stationary (resp. variable bandwidth) Gabor frames, which give a way of tuning the time (resp. frequency) resolution as a function of the time (resp. frequency) variable (see [19], [20] for details). Non-stationary Gabor frames are constructed by varying the window length as a function of the time variable n.

C. Frame Multipliers

Let us consider a pair of frames (resp. Bessel sequences) $\{g_\lambda, \lambda \in \Lambda\}$ and $\{h_\lambda, \lambda \in \Lambda\}$ in the Hilbert space $\mathcal{H}$. A Frame Multiplier (resp. Bessel Multiplier [21]) is an operator $\mathbb{M}_{\boldsymbol{\sigma};g,h}$ which acts on signals $x \in \mathcal{H}$ by pointwise multiplication in the transform domain with a given symbol denoted by $\boldsymbol{\sigma}$, i.e. a sequence $\boldsymbol{\sigma} = \{\sigma[\lambda], \lambda \in \Lambda\}$. Denoting by $\Upsilon_{\boldsymbol{\sigma}}$ the linear operator of pointwise multiplication with a sequence $\boldsymbol{\sigma}$, we write

$$\mathbb{M}_{\boldsymbol{\sigma};g,h} = \mathcal{V}_h^* \circ \Upsilon_{\boldsymbol{\sigma}} \circ \mathcal{V}_g : \mathcal{H} \to \mathcal{H} \ , \tag{4}$$

where $\circ$ denotes the composition operator. In other words

$$\mathbb{M}_{\boldsymbol{\sigma};g,h} x = \sum_\lambda \boldsymbol{\sigma}[\lambda] \langle x, g_\lambda \rangle h_\lambda \ . \tag{5}$$

$\boldsymbol{\sigma}$ is called the Frame mask (or the upper symbol in the mathematics literature) and can be viewed as a *transfer function* in the considered signal representation domain. $\mathbb{M}_{\boldsymbol{\sigma};g,h}$ is a linear operator on the space of signals and is diagonal in the signal representation G_λ.

When the Fourier basis is used for signal representation, multipliers coincide the standard convolution operator and the symbol associated is the transfer function in the frequency domain of a linear time-invariant system. In the case of a time-frequency representation of signals, multipliers can be used to implement time-varying systems [4], [5]. A more theoretical approach in the case of Gabor Multiplier can be found in [22] in the context of Gabor analysis. It can be shown [23] [24] that underspread operators (i.e. operators that don't involve large time-frequency shifts) can be well approximated by Gabor multipliers, provided the window is suitably chosen.

In the following, we shall be concerned below with a different problem, namely the problem of estimating the mask of a Gabor multiplier, given input and output signals. The multiplier estimation problem which we shall address below is formulated in the frame language. While only Bessel sequences are needed in what follows, it is generally more convenient in applications to work with frames.

III. Multipliers Estimation

We now address the following problem: given input or output signals, and analysis and synthesis atoms, estimate the transfer function of the optimal multiplier that (approximately) maps the input to the output signals. Here, optimality is defined in terms of minimization of a functional, which takes the form of a regularized quadratic error. The estimation problem is posed as a regularized least square problem, which is reformulated as a linear inverse problem. We also address at the end of this section some generalizations, including multiple multipliers, i.e. linear combinations of multipliers.

A. Formulation of the problem

Let $x_0^{(k)}$ and $x_1^{(k)}$ denote respectively input and output signals, labelled by $k = 1, .., \kappa$. We assume the following model

$$x_1^{(k)} = \mathbb{M}_{\boldsymbol{\sigma};g,h} x_0^{(k)} + \epsilon^{(k)} \ ,$$

where the $\epsilon^{(k)}$ represent perturbations, modeled as independent realizations of an additive Gaussian noise, and $\boldsymbol{\sigma}$ is an unknown Gabor mask, which we aim at estimating, and g and h are dual windows. A possible solution is obviously $\boldsymbol{\sigma} = \sum_k X_1^{(k)} / \sum_k X_0^{(k)}$, where $X_i^{(k)}$ denote the Gabor transform of $x_i^{(k)}$ with window g. However, such a solution is not bounded in general, because nothing prevents the denominator from vanishing or becoming very small. In worst cases, for example when the source and the target are two pure-tones signals with different frequencies, such a Gabor mask doesn't exist. Then, seeking for a Gabor mask as the solution of a regularized inverse problem provides the existence and the uniqueness of the solution (assuming the regularization term d (see below) is a convex function). More precisely, we seek $\boldsymbol{\sigma} \in \mathbb{C}^{M \times N}$ which minimizes the expression

$$\Phi[\boldsymbol{\sigma}] = \sum_{k=1}^{\kappa} \| x_1^{(k)} - \mathbb{M}_{\boldsymbol{\sigma};g,h} x_0^{(k)} \|^2 + \mu \ d[\boldsymbol{\sigma}], \tag{6}$$

where $d[\boldsymbol{\sigma}]$ is a regularization term, whose influence on solution is controlled by the parameter μ. The role of μ is to control the balance between the reconstruction properties of the Gabor mask and the regularization function which adds an *a priori* knowledge of the solution.

Noticing that given x_0, $\mathbb{M}_{\boldsymbol{\sigma};g,h} x_0$ can be seen as a linear operator acting on $\boldsymbol{\sigma}$, we introduce the linear operators $A^{(k)}$ defined by

$$A^{(k)} = \mathcal{V}_h^* \circ \Upsilon_{X_0^{(k)}} : \ \boldsymbol{\sigma} \to A^{(k)} \boldsymbol{\sigma} = \mathcal{V}_h^* \left(\boldsymbol{\sigma} \mathcal{V}_g x_0^{(k)} \right) \ , \tag{7}$$

Then we have

$$\mathbb{M}_{\boldsymbol{\sigma};g,h} x_0^{(k)} = A^{(k)} \boldsymbol{\sigma} \ ,$$

which leads to the following reformulation of the problem. The adjoint of the $A^{(k)}$ reads

$$A^{(k)^*} = \Upsilon_{\overline{X}_0^{(k)}} \circ \mathcal{V}_h \ . \tag{8}$$

For classical choices of d (more precisely, for d with invertible gradient), the solutions of the minimization problem leads to a (generally non-linear) problem. For example, in the finite-dimensional situation $\mathcal{H} = \mathbb{C}^L$, if $d(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma}\|_2^2$, then

$$\underline{\boldsymbol{A}} = \begin{bmatrix} A^{(1)} \\ \cdots \\ A^{(\kappa)} \end{bmatrix} \quad \text{and} \quad \underline{\boldsymbol{A}}^* = \begin{bmatrix} A^{(1)^*} & \cdots & A^{(\kappa)^*} \end{bmatrix}$$

Figure 1: Concatenation of matrices for the inverse problem with multiple input and multiple output

$\nabla_{\overline{\boldsymbol{\sigma}}} d(\boldsymbol{\sigma}) = \boldsymbol{\sigma}$, where $\overline{\boldsymbol{\sigma}}$ denotes the complex conjugate of $\boldsymbol{\sigma}$. The problem becomes also linear and $\boldsymbol{\sigma}$ is the solution of the (generally huge) matrix problem

$$\left(\sum_k A^{(k)^*} A^{(k)} + \mu I \right) \boldsymbol{\sigma} = \sum_k A^{(k)*} x_1^{(k)} \tag{9}$$

We now limit to the finite-dimensional situation, and turn to matrix notations. Organizing the signals $x^{(k)} \in \mathbb{C}^L$ as column vectors, we formulate the estimation problem as an inverse problem. Introduce the vectors $\underline{\boldsymbol{X}}_1 \in \mathbb{C}^{\kappa L}$ defined by vertical concatenation of signals $x_1^{(k)}$, and matrices $\underline{\boldsymbol{A}} \in \mathbb{C}^{\kappa L \times MN}$ defined by vertical concatenation of matrices $A^{(k)}$ as illustrated on Figure 1, equation (6) can be rephrased as

$$\Phi[\boldsymbol{\sigma}] = \|\underline{\boldsymbol{X}}_1 - \underline{\boldsymbol{A}}\boldsymbol{\sigma}\|_2^2 + \mu \; d(\boldsymbol{\sigma}) \; . \tag{10}$$

Notice that the operator $\underline{\boldsymbol{A}}$ depends on the source signals. As illustrated on Figure 1, its adjoint is given by horizontal concatenation of the adjoint operators $A^{(k)^*}$ defined in (8). Even in simple situations like $d(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma}\|_2^2$, where a closed form expression for the solution of (10) exists, the latter can hardly be exploited practically, as the problem (9) involves huge matrix calculus. To fix the ideas, for $\kappa = 8$ signals with $L = 2^{15}$ (approximately 727 msec at sampling rate $f_s = 44100$ Hz), and Gabor frame parameters chosen $M = 1024$ and $a = 128$, the size of matrix $\underline{\boldsymbol{A}}$ is $\kappa L \times MN \approx 2^{18} \times 2^{18}$. In such cases, as well as cases where no closed form solution exist, we rather rely on dedicated numerical algorithms.

A quick approximate solution of the inverse problem can also be obtained by replacing the $A^{(k)^*} A^{(k)}$ by their diagonal (see section III-B below). This can lead to simple closed-form solutions which may sometimes be of acceptable quality in experiments on audio signals, and easy to evaluate. However, we here propose to use iterative algorithms [25], [26] that converge to the exact solution of the initial problem.

B. The diagonal approximation

When the dictionary differs from a orthonormal basis, the formulation (10) involves non diagonal matrices $A^{(k)^*} A^{(k)}$, where the non diagonal terms arise from the correlations between the atoms of the representation. A first approach [7] is to formulate the problem directly in the transform domain, which amounts to a reduction of the $A^{(k)}$ to their diagonal :

$$\tilde{\Phi}[\boldsymbol{\sigma}] = \sum_k \|X_1^{(k)} - X_0^{(k)} \boldsymbol{\sigma}\|_2^2 + \mu \; d(\boldsymbol{\sigma}), \tag{11}$$

For well chosen d, explicit solutions exist. For example, for the quadratic regularization $d(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma} - \boldsymbol{\sigma}^{(r)}\|_2^2$ and some fixed reference mask $\boldsymbol{\sigma}^{(r)}$, the regularized solution resulting from the corresponding variational equations reads

$$\tilde{\boldsymbol{\sigma}} = \frac{\sum_k \overline{X_0^{(k)}} X_1^{(k)} + \mu \boldsymbol{\sigma}^{(r)}}{\sum_k |X_0^{(k)}|^2 + \mu} \; . \tag{12}$$

When a tight frame with frame bounds $A = B = 1$ is used and the reference mask is chosen as $\boldsymbol{\sigma}^{(r)} = 1$, the penalizations for $\boldsymbol{\sigma}$ tend to favor Gabor multipliers close to the identity operator. Similarly, for a frame and its dual, multipliers closed to the identity are favored when $\boldsymbol{\sigma}^{(r)} = 1$.

Other choices for the regularization function will be guided by the applications. For example, a regularization function $d(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma} - \boldsymbol{\sigma}^{(r)}\|_1$ will promote sparsity between the coefficients of the solution [27]. This choice leads to a solution

$$\tilde{\boldsymbol{\sigma}} = \begin{cases} \frac{|X_0||X_1 - \boldsymbol{\sigma}^{(r)} X_0| - \mu/2}{|X_0|^2} e^{i\varphi_{\boldsymbol{\sigma}}} + \boldsymbol{\sigma}^{(r)} & \text{if } \; |X_0||X_1 - \boldsymbol{\sigma}^{(r)} X_0| \geqslant \mu/2 \\ \boldsymbol{\sigma}^{(r)} & \text{else} \end{cases}$$

where we choose $\kappa = 1$ for the sake of simplicity and the phase of the Gabor mask is given by $\varphi_{\boldsymbol{\sigma}} = \arg\left(\overline{X_0}(X_1 - \boldsymbol{\sigma}^{(r)} X_0\right)$.

C. Iterative shrinkage algorithms

We present in this section a general method to estimate the mask of a multiplier, given input and output signals. The formulation given in (10) for our problem, together with the choice of regularization allows us to use iterated shrinkage algorithms similar to those described in [28], [29] to which we refer for more details and proofs. Those algorithms can also be formulated in the language of proximal algorithms (see [26] for a review), but we limit the discussion here to Landweber-type approaches.

Our problem, as explained previously, is written as follows. We consider a tight frame (we shall see later on that a Bessel sequence would be enough for the convergence), and seek solutions of

$$\min_{\boldsymbol{\sigma}} \; \Phi(\boldsymbol{\sigma}) \; , \quad \text{with} \quad \Phi(\boldsymbol{\sigma}) = \|\underline{\boldsymbol{X}}_1 - \underline{\boldsymbol{A}}\boldsymbol{\sigma}\|^2 + \mu d(\boldsymbol{\sigma}) \tag{13}$$

It is known that for $d(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma}\|_p^p$ with $p \geqslant 1$, this functional is convex and then has a unique minimizer. However, the latter is generally hard to compute in large dimensions, and one has to resort to appropriate numerical algorithms. The solution that was proposed in [28], which converges to the solution with minimal assumptions on $\underline{\boldsymbol{A}}$, is based upon surrogate functionals. Assuming $\underline{\boldsymbol{A}}$ is bounded, we can pick a constant C such that $\|\underline{\boldsymbol{A}}^* \underline{\boldsymbol{A}}\|_{Op} < C$ (with $\|-\|_{Op}$ the operator norm, here the largest singular value of $\underline{\boldsymbol{A}}$). In the considered situation, $\|\underline{\boldsymbol{A}}^* \underline{\boldsymbol{A}}\|_{Op}$ can be estimated explicitly and reads

$$\|\underline{\boldsymbol{A}}^* \underline{\boldsymbol{A}}\|_{Op} \leqslant \frac{\kappa}{A} \max_k \|X_0^{(k)}\|_\infty^2 \tag{14}$$

where $\|-\|_\infty$ gives the maximum value of the modulus of the frame coefficients. The details of this calculation have been reported in the appendix. The evaluation of $\|\underline{\boldsymbol{A}}^* \underline{\boldsymbol{A}}\|_{Op}$ can also be done numerically using a power iteration algorithm as proposed in [30].

Remark 1: Following [31], the derivation of real valued functions of several complex variables are made with real data methods by considering $(\boldsymbol{\sigma}, \overline{\boldsymbol{\sigma}})$ as two independent variables. In this context, the gradient are evaluated with respect to the variable $\overline{\boldsymbol{\sigma}}$.

Given these notations, fix $\boldsymbol{\alpha} \in \mathbb{C}^{MN}$, and introduce the surrogate functional

$$\Phi^{SUR}(\boldsymbol{\sigma}; \boldsymbol{\alpha}) = \Phi(\boldsymbol{\sigma}) - \|\underline{\boldsymbol{A}}\boldsymbol{\sigma} - \underline{\boldsymbol{A}}\boldsymbol{\alpha}\|_2^2 + C\|\boldsymbol{\sigma} - \boldsymbol{\alpha}\|_2^2 \ . \quad (15)$$

The latter is still convex for any $\boldsymbol{\alpha} \in \mathbb{C}^{MN}$, and has the advantage to admit a closed form expression for its unique minimizer. Starting from some initial guess $\boldsymbol{\alpha} = \boldsymbol{\sigma}_0 \in \mathbb{C}^{M\times N}$, the idea is then to iteratively determine the minimizer of (15) for $\boldsymbol{\alpha} = \boldsymbol{\sigma}_{n-1}$. This therefore yields the iterative method summarized in **Algorithm** 1.

Algorithm 1 ISTA

Require: $\boldsymbol{\sigma}_0$
  while The relative error is bigger than a threshold : **do**
    $\boldsymbol{\sigma}_n = \text{argmin}_{\boldsymbol{\sigma}}\{\Phi^{SUR}(\boldsymbol{\sigma}; \boldsymbol{\sigma}_{n-1}), \quad \boldsymbol{\sigma} \in \mathbb{C}^{M\times N}\}$
  end while

For the sake of clarity let us set the gradient scheme applied on the first term of our problem (10)

$$\boldsymbol{\beta}_{n-1} = \boldsymbol{\sigma}_{n-1} - \frac{1}{C}\underline{\boldsymbol{A}}^*(\underline{\boldsymbol{X}}_1 - \underline{\boldsymbol{A}}\boldsymbol{\sigma}_{n-1}) \ ,$$

where $1/C$ plays the role of a gradient step. The following choices for the regularization terms are of interest.

- $d(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma} - \boldsymbol{\sigma}^{(r)}\|_2^2$. This choice leads to a simple gradient algorithm, which is a damped version of Landweber iterative method (corresponding to the case $\mu = 0$) and expressed as

  $$\boldsymbol{\sigma}_0 \in \mathbb{C}^{M\times N}, \qquad \boldsymbol{\sigma}_n = \frac{\boldsymbol{\beta}_{n-1} + \boldsymbol{\sigma}^{(r)}\mu/C}{1 + \mu/C}$$

  This algorithm is also called an iterative shrinkage algorithm as it adds some weighted adjustments on the gradient to $\boldsymbol{\beta}_{n-1}$. The case $\boldsymbol{\sigma}^{(r)} = 0$ corresponds to the Tikhonov regularization.
- $d(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma} - \boldsymbol{\sigma}^{(r)}\|_1$. This choice leads to a thresholding iterative algorithm expressed as

  $$\boldsymbol{\sigma}_0 \in \mathbb{C}^{M\times N}, \qquad \boldsymbol{\sigma}_n = \mathcal{S}_{\mu/\mathcal{C}}(\boldsymbol{\beta}_n - \boldsymbol{\sigma}^{(r)}) + \boldsymbol{\sigma}^{(r)}$$

  where $\mathcal{S}_\mu$ is a thresholding operator defined componentwise: for $\boldsymbol{\beta} = \{\beta[m,n]\} \in \mathbb{C}^{MN}$, $(\mathcal{S}_\mu(\boldsymbol{\beta}))[m,n] = S_\mu(\beta[m,n])$, with

  $$\mathcal{S}_\mu(z) = \begin{cases} (|z| - \mu)\frac{e^{i\arg(z)}}{2} & \text{if } |z| > \mu \\ 0 & \text{if } 0 \leqslant |z| \leqslant \mu \end{cases}$$

Technically speaking, the algorithms described above belong to the class of first order methods and therefore converge as $O(1/n)$. The authors in [32] proposed a second order algorithm that converge as $O(1/n^2)$ without important increased complexity in the iterations. This approach is outlined in **Algorithm** 2. Numerical results are displayed in Figure 3, and discussed in Section IV below.

Algorithm 2 FISTA

Require: $\boldsymbol{s}_1 = \boldsymbol{\sigma}_0, t_1 = 1$
  while The relative error is bigger than a threshold : **do**
    $\boldsymbol{\sigma}_n = \text{argmin}\{\Phi^{SUR}(\boldsymbol{s}; \boldsymbol{s}_n) \ : \ \boldsymbol{s} \in \mathbb{C}^{M\times N}\}$
    $t_{n+1} = \frac{1+\sqrt{1+4t_n^2}}{2}$
    $\boldsymbol{s}_{n+1} = \boldsymbol{\sigma}_n + \frac{t_n - 1}{t_{n+1}}(\boldsymbol{\sigma}_n - \boldsymbol{\sigma}_{n-1})$
  end while

Notice that FISTA does not ensure monotone decay of the objective function. In situations where this behavior is observed, a monotone version of FISTA called MFISTA, proposed in [26], can be used.

As we will see with audio examples in section IV, to avoid creating phase distortions, it is also interesting to consider other cases like

$$d(\boldsymbol{\sigma}) = \||\boldsymbol{\sigma}| - 1\|_2^2 \ , \quad (16)$$

or

$$d(\boldsymbol{\sigma}) = \||\boldsymbol{\sigma}| - \log|\boldsymbol{\sigma}| - 1\|_1 \ .$$

In such cases, the usual convergence analysis unfortunately does not apply straightforwardly. Indeed, these regularization terms $d(\boldsymbol{\sigma})$ are not convex and the uniqueness of the solution in general situations is lost. However, these penalization terms are actually convex with respect to the modulus of $\boldsymbol{\sigma}$, which make us believe that convergence could be provable in the context of proximal algorithm where the uniqueness of the proximal gradient is preserved. Furthermore, it can be shown in such cases [26] that the cost function decreases at each iteration of ISTA. Numerical experiments presented hereafter on Figure 3 illustrate the behavior of the cost function for both ISTA and FISTA in the convex and the non convex case.

Independently of convergence considerations, the blind application of the above approach to the situation (16) yields the following update rule : given some initialization $\boldsymbol{\sigma}_0 \in \mathbb{C}^{M\times N}$, iterate

$$\boldsymbol{\sigma}_n = \frac{|\boldsymbol{\beta}_{n-1}| + \mu/C}{1 + \mu/C} \ e^{i\arg(\boldsymbol{\beta}_{n-1})} \ .$$

The algorithm itself is still easily implemented, and shows good experimental convergence properties when suitably initialized. In particular, for audio applications, this approach has the advantage of avoiding artifacts caused by an inaccurate phase estimation for large values of μ. More details about convergence and behavior of the solutions in such cases are provided below in audio examples.

D. Gabor Multipliers

We have shown above a general situation for the frame multiplier estimation problem, combined with the use of efficient algorithms. In the following, we will also concentrate on Gabor frames, which are the most relevant framework to deal with audio signals in the context of Analysis/Transformation/Synthesis. In addition, the general estimation problem we present in the section III-A, III-B and III-C can be applied with any frame and Bessel sequence.

When Gabor frames are used to represent signals, the corresponding multipliers (called Gabor Multipliers) have been studied by several authors (see [22], [23] and references therein). A Gabor Multiplier (GM for short) $\mathbb{M}_{\boldsymbol{\sigma};g,h} : x \to \mathbb{M}_{\boldsymbol{\sigma}} x$ is defined by

$$\mathbb{M}_{\boldsymbol{\sigma};g,h} x = \sum_{m,n} \boldsymbol{\sigma}[m,n] \mathcal{V}_g x[m,n] h_{mn} \ . \tag{17}$$

$\boldsymbol{\sigma}$ is called *Gabor mask* (or the upper symbol in the mathematics literature) and can be viewed as a *time-frequency transfer function* (so that $\mathbb{M}_{\boldsymbol{\sigma}}$ is seen as a time-varying filter). Then, a Gabor Multiplier acts on a representation and influence the relation between the coefficients of the representation.

Gabor coefficients provide a time-frequency representation from which synthesis can be done efficiently. It is a standard practice in many aplications to perform partial resynthesis from a subset of Gabor coefficients [6], [33], and we would like to point out that such actions actually correspond to simple instances of Gabor multipliers, using a binary mask $\boldsymbol{\sigma}$. Estimation of a Gabor mask from input and output signals can also be seen as a particular case of offline system identification, performed directly in the time-frequency domain, in the spirit of [9], [10]. The type of applications we shall be interested in section IV below are more concerned with audio signal categorization, timbre characterization and sound morphing. For the first two applications, our main point is that the estimated masks contain highly relevant information, that can be used for categorization and characterization. The morphing application uses the synthesis capability of Gabor frames.

E. Generalizations

The simple "input-output" model $x_0^{(k)} \to x_1^{(k)} = \mathbb{M}_{\boldsymbol{\sigma}} x_0^{(k)} + \epsilon^{(k)}$ is not always accurate enough to properly characterize the differences between two signal classes x_0 and x_1. Further transformations, such as time and/or frequency shifts, are poorly described by time-frequency multipliers. We now show that the generic scheme described above can also handle more complex situations.

- *Multiple Gabor Multipliers (MGM).* Given an analysis window g, a Multiple Gabor Multiplier (MGM) is defined as a linear combination of Gabor multipliers using different synthesis windows:

  $$\mathbb{M}_{\underline{\boldsymbol{\sigma}}} x = \sum_{\lambda \in \Lambda} \sum_{m,n} \boldsymbol{\sigma}_\lambda[m,n] \langle x, g_{m,n} \rangle (h_\lambda)_{m,n}$$

  where $h_\lambda, \lambda \in \Lambda$ is a family of synthesis windows, and $\underline{\boldsymbol{\sigma}}$ is a corresponding family of masks. In order to ensure that such a MGM be well defined, one can for instance assume that $\sum_\lambda \sup_{m,n} |\boldsymbol{\sigma}_\lambda[m,n]| < \infty$ and $\max_\lambda \|h_\lambda\|_2 < \infty$. In particular, the synthesis windows can be chosen as a family of time-frequency shifted copies of a given synthesis window h, by taking $\lambda = (k,\ell)$ and $h_{k,\ell}[n] = e^{2i\pi\ell\nu_1[n-kb_1]} h[n-kb_1]$, for some lattice constants b_1, ν_1. More generally [34], for any frames, Multiple Multipliers can well approximate any bounded operators.

  Given a pair of input and output signals (x_0, x_1) as before the corresponding estimation problem

  $$\min_{\boldsymbol{\sigma}_\lambda, \lambda \in \Lambda} \left[\frac{1}{2} \left\| x_1 - \sum_{\lambda \in \Lambda} \mathbb{M}_{\boldsymbol{\sigma}_\lambda} x_0 \right\|^2 + d[\boldsymbol{\sigma}_\lambda] \right]$$

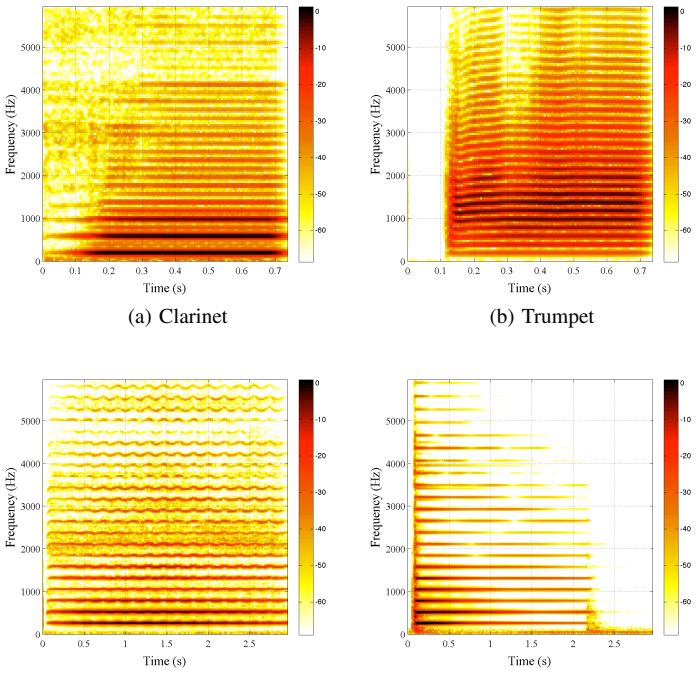
  can again be rephrased as a linear inverse problem by a suitable rearrangement of the terms. However, much care is required in this context for designing the regularization terms $d[\boldsymbol{\sigma}_\lambda]$. We will not discuss further these issues in the current paper.
- *GM perturbation of a stationary system.* Gabor multipliers are essentially local transformations in the time-frequency domain. In some situations, the transformation can be modelled as the composition of a stationary (translation invariant) system and such a local transformation. The model thus becomes

  $$x_1^{(k)} = F^{(k)} \mathbb{M}_{\boldsymbol{\sigma}} x_0^{(k)} + \epsilon^{(k)} \ ,$$

  where $F^{(k)}$ represent known convolution operators, which have been applied on each source signal. The algorithms described above can again be adapted to such a situation, and the matrix $\underline{\boldsymbol{A}}$ now depends both on the source $x_0^{(k)}$ and the filters $F^{(k)}$.

IV. Computational experiments

We show in this section numerical experiments that illustrate the convergence of the proposed algorithms, in the context of time-frequency analysis of audio signals with Gabor frame. We also give a few examples of Gabor masks estimated from real audio signals, in order to discuss the role of the regularization term and the estimation method. Finally, we discuss in more details potential applications of Gabor masks between audio signals. We consider a clarinet and a trumpet from the IOWA database [35] of $L = 2^{15}$ samples, with fundamental frequency f_0 =196 Hz (G3). We also consider violin and piano sound signals of from the RWC database [36] of $L = 2^{17}$ samples, with fundamental frequency $f_0 = 261$ Hz (C4). Their time-frequency representations are shown in Figure 2. Clarinet and trumpet signals were obtained using a tight Gabor frame, obtained from an Hann window and parameter values $M = 1024$, $a = 256$. Violin and piano signals were obtained using a tight Gabor frame, obtained from an Hann window and parameter values $M = 2048$, $a = 512$. Here, we also consider $h = g$ and a tight Gabor frames with frame bounds $A = B = 1$. These sounds show significant differences, which can be interpreted physically, and which will be captured by the estimated Gabor masks. All signals exhibit a harmonic structure, with the following most striking differences, regarding their time-frequency behavior (time evolution of the harmonics, duration of the attack, frequency repartition, ...). We also estimate Gabor masks between the clarinet and the trumpet and between the piano and the violin, with the different configurations of the problem explained in section III, in order to reveal the role of the regularization function and the role of the Gabor mask estimation method. On all figures, amplitudes are represented with a logarithmic scale.



(a) Clarinet (b) Trumpet

(c) Violin (d) Piano

Figure 2: Spectrograms of audio signals used in the experiment

Remark 2: Gabor mask will be sensible to small time-shifts between source and target signals. Prior to mask estimation, the signals have therefore been readjust by hand, such that their onsets coincide.

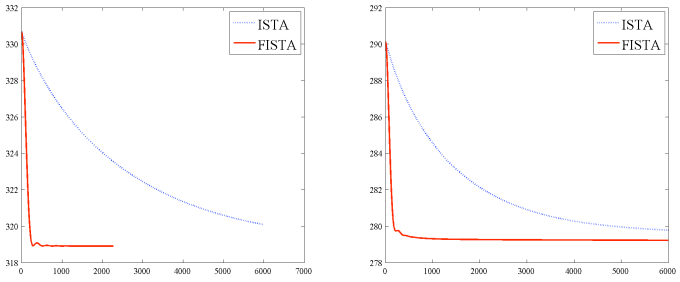
Sounds examples and `Matlab` code can be found on `www.lma.cnrs-mrs.fr/~kronland/olivero/ieee2012/ieee2012.html`

A. Audio applications: Sounds timbre transformation

In the context of audio signals analysis, a lot of applications involve the evaluation of a time-varying transfer function between two signals. The information present in the Gabor masks was shown in [12], [13] to be relevant for audio signals analysis, in a categorization context. Dissimilarity measures extracted directly from masks were shown to yield good classifications of single note signals from four different classes of musical instruments, which proves that the Gabor mask captures a sensible information in the time-frequency domain.

From a synthesis point of view, the transfer functions obtained as proposed here can be used for performing sound morphing between a pair of sounds, or more generally for sound synthesis. Sound morphing covers a wide variety of techniques whose aim is to "interpolate" between two sound signals, with perceptually relevant characteristics. A detailed review of most of existing approaches can be found in [37]. We presented in [8], [14] a morphing strategy based on Gabor Multipliers. Our approach exploits a Gabor representation as low level representation, from which Gabor masks are estimated as above. The regularization parameter μ is then used as a control (tuning) parameter, allowing interpolation between source (large μ) and target (small μ). We also show in [14] that, with the regularization function $d(\boldsymbol{\sigma}) = \||\boldsymbol{\sigma}| - 1\|_2^2$, the set of μ values used to construct a sound morphing clearly depends on the energy contained in the sounds used as source and target of the morphing process. Small and large value of μ can also be estimated from the source and target signals.

B. Behavior of the algorithms



(a) Convex (b) Non convex

Figure 3: Convergence of ISTA and FISTA: objective function (see (10)) as a function of the iteration index, for $\kappa = 1$

We used the algorithms described above from the clarinet to the trumpet signal. The algorithm convergence is illustrated in Figure 3, where the cost function is plotted as a function of the number of iterations, for both ISTA and FISTA. For that particular experiment, we set $\mu = 10^{-3}$ and the regularization term was $d(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma} - 1\|_2^2$. It is known that such algorithms are faster for large value of μ and benefit from the use of a good initialization. The latter can be chosen as the explicit solution of the approximation problem given by equation (11), or as a (converged) solution obtained for a bigger value of μ. We choose the one which gives the lower value of the current cost function, over the set of Gabor masks evaluated previously by the algorithm with lower values for μ and evaluated by diagonal approximation with the current value for μ. We observe in practice a similar behavior when other values for the parameter μ are chosen.

Remark 3: The multiplier estimated by our approach are far from being invertible [38], as they intrinsically depend on the sounds used as input and output signals. Then, inverting the role of the input and output signals will lead to different sounds.

C. Qualitative analysis of the estimated Gabor masks : influence of the regularization function

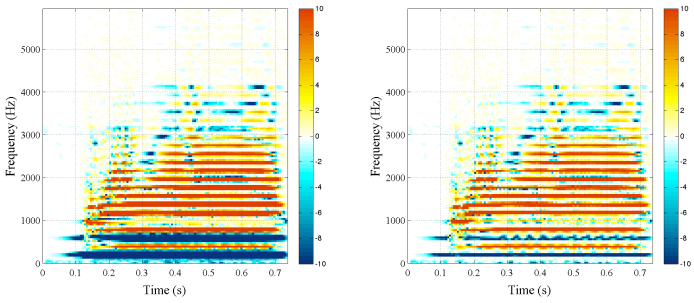
We used the algorithms described above from the clarinet to the trumpet signal. First, let us compare the iterative methods with the diagonal approximations, using the convex regularization term $d(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma} - 1\|_2^2$, and a moderate value of the regularization parameter μ, which leads to in "intermediate" sound whose timbre can be heard between the timber of the source and target sounds. These two estimated Gabor masks are presented in Figure 4, for $\mu = 10^{-3}$. The comparison shows that the iterative method tends to provide clearer harmonic

components for the Gabor mask. The increased computational cost induced by the iterative approach is therefore justified.

However, a closer examination reveals the presence of spurious oscillations in the estimated Gabor mask. These oscillations turn out to result from the inappropriate choice of the regularization term $d(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma} - 1\|^2$. The latter constrains the argument of the mask and therefore does not account properly for relative phase behaviors of input and output signals. For the diagonal approximation (see equation (12) with $\kappa = 1$ and $\boldsymbol{\sigma}^{(r)} = 1$), we can show how these phase artefacts appear in the Gabor mask, when a time-frequency point $[m, n]$ present a phase difference equal to π between the source $X_0[m, n]$ and target $X_1[m, n]$, and when $|X_0[m, n]|^2 = \mu$, the Gabor mask at this point is given by

$$\tilde{\boldsymbol{\sigma}} = \frac{\overline{X_0} X_1 + \mu}{|X_0|^2 + \mu} = \frac{|X_0|^2 e^{i\pi} + \mu}{|X_0|^2 + \mu} = \frac{e^{i\pi} + 1}{2} = 0$$

where we omitted the index $[m, n]$ for the sake of clarity. At this time-frequency point, the Gabor mask vanishes. As μ is a global value applied on all the time-frequency coefficients, the presence of zeros in the Gabor mask artificially creates amplitude modulation in the Gabor mask as observe in Figure 4.

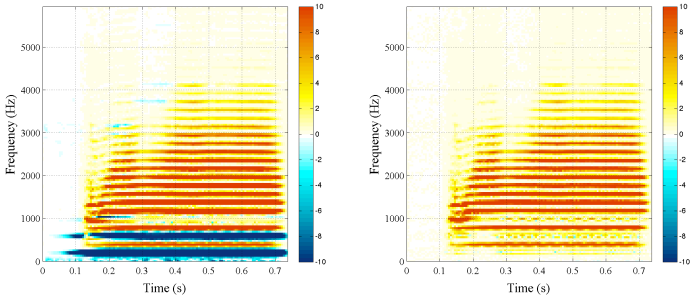


(a) Diagonal approximation (b) FISTA

Figure 4: Gabor masks obtained for $\mu = 10^{-3}$, $d = \|\boldsymbol{\sigma} - 1\|_2^2$.

This motivated us to turn to the non convex constraint $d(\boldsymbol{\sigma}) = \||\boldsymbol{\sigma}| - 1\|_2^2$. These two estimated Gabor masks are presented in Figure 4, for $\mu = 10^{-3}$, where we can clearly see that the artificial amplitude modulation are not present any more in the Gabor mask estimated by diagonal approximation and FISTA. This behavior of the Gabor mask clearly shows the importance of the phase in the Gabor domain, and how a bad estimation of the Gabor mask phase can provide spurious artifacts. Again, the solution found by the algorithm is quite different from the diagonal approximation, and significantly sparser. Finally, the choice $d(\boldsymbol{\sigma}) = \||\boldsymbol{\sigma}| - 1\|_2^2$ clearly outperforms $d(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma} - 1\|_2^2$, and we can reasonably argue that regularization functions acting on the modulus of the Gabor masks should to be systematically used with audio data.

The morphed sounds obtained using FISTA and the diagonal approximation appear to be quite close to each other in this example, as confirmed by informal listening tests to the corresponding sounds. This originates from the fact that, even though the masks are significantly different, the synthesis operator compensates from these differences. Other sounds examples have been reported on the url linked to the paper.



(a) Diagonal approximation (b) FISTA

Figure 5: Gabor masks obtained for $\mu = 10^{-3}$, $d = \||\boldsymbol{\sigma}| - 1\|_2^2$.

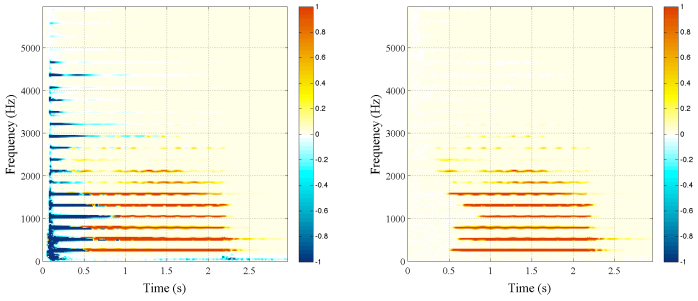
D. Qualitative analysis of the estimated Gabor masks : influence of the estimation method

We estimate here the Gabor masks from the violin to the piano. This case is a more difficult one, as the two sounds features different harmonic structure across time. The high frequencies in the piano damp faster than those of the violin, which features frequency modulation caused by the vibrato. Estimated Gabor masks are shown in Figure 6 and the reconstructed signals are shown in Figure 7. We first notice that the estimated Gabor masks behave differently with regards to the estimation method. The Gabor mask modulus obtained with FISTA tends to be closer to 1 during the first 0.5 seconds. However, it is interesting to notice that the Gabor mak obtained by FISTA generates a signal with a smoother attack, although the modulus of the Gabor mask equals to 1 during the attack. Once again, this experiment shows the importance of the Gabor mask phase, which is responsible here for the transformation of the attack between these two particular sounds. In addition, we can hear that sounds generated using the diagonal approximation are of low perceptual quality : they feature unexpected artifacts, and are hardly perceived as "intermediate sounds" between violin and piano. When morphed sounds are generated using FISTA, the perceived quality improves significantly.

These preliminary results, to be confirmed by more systematic perceptive evaluations, tend to show that the iterative method combined with a penalization of the mask modulus, can provide a valuable approach for the considered sounds morphing problem.

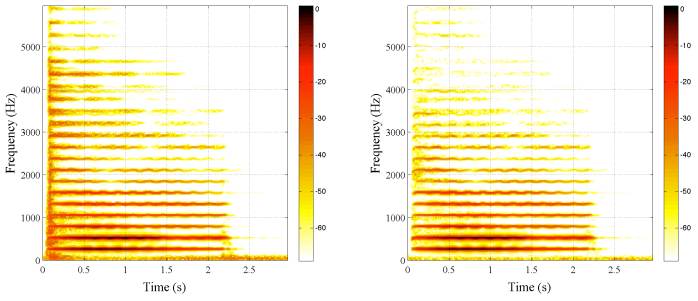
V. Conclusion and Perspectives

In this paper, we have proposed and developed an iterative scheme for the estimation of a linear transformation modelled as a (multiple) frame multiplier, and illustrated it with a few examples in the context of audio signal processing. While multiplier estimation can be performed by simpler techniques (such as the diagonal approximation described in Section III-B) our results show that estimates provided by our approach yield transfer functions with a neater structure. This turns out to be relevant in morphing applications, in situations where the input and output signals are different enough. These findings remain to be confirmed by more thorough experiments



(a) Diagonal approximation (b) FISTA

Figure 6: Gabor masks obtained for $\mu = 10^{-2}$, $d = \||\boldsymbol{\sigma}|-1\|_2^2$.



(a) Diagonal approximation (b) FISTA

Figure 7: Reconstructed signals obtained for $\mu = 10^{-2}$, $d = \||\boldsymbol{\sigma}| - 1\|_2^2$.

using larger sound databases. In addition, our approach is easily extended to more complex settings such as multiple multipliers.

This work is a part of a program which aims at exploiting Gabor multipliers and related techniques for audio signal analysis, including categorization problems, and sound morphing. In this respect, we plan to investigate further extensions of this work including constrained versions of the estimation problem we have been considering here. Questions regarding convergence properties of our scheme in situations where non convex regularization terms are used are also of interest. Applications to audio morphing, in the spirit of our previous work in [8] are under progress.

ACKNOWLEDGEMENTS

This work was supported in parts by the WWTF project MULAC (Frame Multipliers: Theory and Application in Acoustics; MA07-025) and the ANR project METASON (CONTINT 2010 ANR-10-CORD-010).

APPENDIX

Proof of the equation (14) :

$$
\begin{aligned}
&\|\underline{\boldsymbol{A}}^*\underline{\boldsymbol{A}}\boldsymbol{\sigma}\|_2^2 \\
&\quad= \left\|\sum_k A^{(k)^*}A^{(k)}\boldsymbol{\sigma}\right\|_2^2 \\
&\quad= \sum_{m,n}\left|\sum_k \overline{X_0^{(k)}[m,n]}\cdot \mathcal{V}_h\mathcal{V}_h^*(X_0^{(k)}\boldsymbol{\sigma})[m,n]\right|^2 \\
&\quad\leqslant \sum_{m,n}\left(\textstyle\sum_k |X_0^{(k)}[m,n]|^2\right)\cdot\left(\sum_k\left|\mathcal{V}_h\mathcal{V}_h^*(X_0^{(k)}\boldsymbol{\sigma})[m,n]\right|^2\right) \\
&\quad\leqslant \kappa\ \max_k\|X_0^{(k)}\|_\infty^2\cdot\sum_{m,n}\sum_k\left|\mathcal{V}_h\mathcal{V}_h^*(X_0^{(k)}\boldsymbol{\sigma})[m,n]\right|^2 \\
&\quad= \kappa\ \max_k\|X_0^{(k)}\|_\infty^2\cdot\sum_k\|\mathcal{V}_h\mathcal{V}_h^*(X_0^{(k)}\boldsymbol{\sigma})\|_2^2 \\
&\quad\leqslant \kappa\ \max_k\|X_0^{(k)}\|_\infty^2\cdot\tfrac{1}{A^2}\sum_k\|X_0^{(k)}\boldsymbol{\sigma}\|_2^2 \\
&\quad\leqslant \tfrac{\kappa^2}{A^2}\max_k\|X_0^{(k)}\|_\infty^4\|\boldsymbol{\sigma}\|_2^2
\end{aligned}
$$

REFERENCES

[1] X. Serra and J. O. Smith, "Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition," *Computer Music Journal*, vol. 14, no. 4, pp. 12–24, 1990.

[2] R. McAulay and T. Quatieri, "Speech analysis/synthesis based on a deterministic plus stochastic decomposition," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, pp. 744–754, August 1986.

[3] S. Mallat, *A wavelet tour of signal processing: the sparse way*. Academic Press, 2009.

[4] M. Portnoff, "Time-frequency representation of digital signals and systems based on short-time fourier analysis," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-28, no. 1, pp. 55–69, 1980.

[5] G. Matz and F. Hlawatsch, *Linear Time-Frequency Filters: Online Algorithms and Applications*. A. Papandreou-Suppappola, Ed., 2002.

[6] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Transactions on Signal Processing*, vol. 52, pp. 1830–1847, July 2004.

[7] P. Depalle, R. Kronland-Martinet, and B. Torrésani, "Time-frequency multipliers for sound synthesis," in *Proceedings of the Wavelet XII conference, SPIE annual Symposium*, pp. 221–224, San Diego, 4-8 September 2007.

[8] A. Olivero, B. Torrésani, and R. Kronland-Martinet, "A new method for Gabor mulipliers estimation : Application to sound morphing," in *EUSIPCO 2010*, Alborg, August 2010.

[9] Y. Avargel and I. Cohen, "On multiplicative transfer function approximation in the short-time fourier domain," *IEEE Signal Processing Letters*, vol. 14, no. 5, pp. 337–340, 2007.

[10] Y. Avargel and I. Cohen, "Adaptive system identification in the short-time fourier domain using cross-multiplicative transfer function approximation," *IEEE Transactions on Audio Speech and Language Processing*, vol. 16, no. 1, pp. 162–173, 2008.

[11] P. Guillemain, C. Vergez, D. Ferrand, and A. Farcy, "An instrumented saxophone mouthpiece and its use to understand how an experienced musician play," *Acta Acustica united with Acustica*, vol. 96, no. 4, pp. 622–634, 2010. OR 21.

[12] A. Olivero, L. Daudet, R. Kronland-Martinet, and B. Torrésani, "Analyse et catégorisation de sons par multiplicateurs temps-fréquence," in *XXIIe colloque GRETSI (Dijon)*, 8-11 septembre 2009. http://documents.irevues.inist.fr/handle/2042/29160.

[13] A. Olivero, "Computing time-frequency maps for timbre discrimination," in *Proceedings of the Digital Audio Effects Conference (Dafx)*, septembre 19-23, 2011.

[14] A. Olivero, P. Depalle, B. Torrésani, and R. Kronland-Martinet, "Sound morphing strategies based on alterations of time-frequency representations by gabor multipliers," in *AES 45th International Conference*, (Helsinki, Finland), 1-4 March 2012.

[15] J. Princen, A. Johnson, and A. Bradley, "Subband/transform coding using filter bank designs based on time domain aliasing cancellation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 12, pp. 2161–2164, 1987.

[16] O. Christensen, *Frames and Bases. An Introductory Course.* Applied and Numerical Harmonic Analysis. Basel Birkhäuser, 2008.

[17] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing (2nd Ed.)*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[18] T. Strohmer, *Gabor Analysis and Algorithms - Theory and Applications*, ch. Numerical Algorithms for Discrete Gabor Expansions, pp. 267–294. Birkhäuser Boston, 1998.
[19] F. Jaillet, *Représentation et traitement temps-fréquence des signaux audionumériques pour des applications de design sonore*. PhD thesis, Université de Provence, 2005.
[20] P. Balazs, M. Dörfler, N. Holighaus, F. Jaillet, and G. Velasco, "Theory, implementation and application of nonstationary Gabor frames," *Journal of Computational and Applied Mathematics*, vol. 236, no. 6, pp. 1481–1496, 2011.
[21] P. Balazs, "Basic definition and properties of Bessel multipliers," *Journal of Mathematical Analysis and Applications*, vol. 325, pp. 571–585, January 2007.
[22] H. G. Feichtinger and K. Nowak, "A first survey of Gabor multipliers," in *Advances in Gabor Analysis. Birkhauser*, pp. 99–128, Birkhäuser, 2002.
[23] M. Dörfler and B. Torrésani, "Representation of operators in the time-frequency domain and generalized Gabor mulipliers," *Journal of Fourier Analysis and Applications*, vol. 16, pp. 261–293, 2010.
[24] M. Dörfler and B. Torrésani, "Approximation of operators by sampling in the time-frequency domain," *Sampling Theory in Signal and Image Processing*, vol. 7, no. 1, pp. 65–76, 2011.
[25] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambribge University Press, New York, 2004.
[26] A. Beck and M. Teboulle, *Gradient-Based Algorithms with Applications to Signal Recovery Problems*. Cambribge University Press, 2010.
[27] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.
[28] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, November 2004.
[29] G. Teschke, "Multi-frame representations in linear inverse problems with mixed multi-constraints," *Applied and Computational Harmonic Analysis*, vol. 22, pp. 43–60, 2007.
[30] M. Kowalski, E. Vincent, and R. Gribonval, "Beyond the narrowband approximation: Wideband convex methods for under-determined reverberant audio source separation," *IEEE Transactions on Audio Speech and Language Processing*, vol. 17(7), Special Issue on: "Processing Reverberant", pp. 1818–1829, September 2010.
[31] T. Adali and S. Haykin, *Adaptive Signal Processing: Next Generation Solutions*. Wiley-IEEE Press, March 2010.
[32] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, March 4, 2009.
[33] U. Kjems, J. B. Boldt, M. S. Pedersen, T. Lunner, and D. Wang, "Role of mask pattern in intelligibility of ideal binary-masked noisy speech," *Journal of the Acoustical Society of America*, vol. 126, pp. 1415–1426, 2009.
[34] P. Balazs, "Matrix-representation of operators using frames," *Sampling Theory in Signal and Image Processing (STSIP)*, vol. 7, pp. 39–54, Jan. 2008.
[35] IOWA, "The iowa music instrument samples," in *http://theremin.music.uiowa.edu*.
[36] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "Rwc music database : Music genre database and musical instrument sound database," in *Proc. of the International Conference on Music Information Retrieval (ISMIR)*, (Baltimore, MD, USA), October 2003.
[37] M. Caetano, *Morphing Isolated Quasi-Harmonic Acoustic Musical Instrument SoundsGuided by Perceptually Motivated Features*. PhD thesis, University of Paris VI - Université Pierre et Marie Curie (UPMC), June 2011.
[38] D. T. Stoeva and P. Balazs, "Invertibility of multipliers," *Applied and Computational Harmonic Analysis*, vol. in press, 2011.

Anaik Olivero received the M.Sc. degree in acoustics from the Université de Provence in 2008, and the Ph.D. degree from Aix-Marseille University in 2012 after completing a thesis on audio signal processing. Currently, she is a post-doctoral researcher at Laboratoire d'Informatique Fondamentale (LIF) in Marseille. Her research interests are in signal processing and machine learning applied to audio signals.

Bruno Torrésani received the PhD degree in mathematical physics from Université d'Aix-Marseille I in 1986, and the habilitation degree from Université d'Aix-Marseille II in 1992. He was researcher at CNRS from 1988 to 1998 at Centre the Physique Théorique, Marseille, France, and is now professor in Mathematics at Université d'Aix-Marseille, and the head of LATP, the mathematics laboratory. He held associate professor positions at Université de Louvain la Neuve, Belgium, University of California at Irvine (USA), and Universidad de La Plata (Argentina). His research interests are mainly in mathematical signal processing, including applied harmonic analysis and functional analysis, probabilistic modeling and statistics, and applications to various domains such as audio signal processing, genomics and neurosciences. and machine learning applied to audio signals.

Richard Kronland-Martinet received a Master degree in theoretical physics in 1980 and a Ph.D. degree in acoustics in 1983 from the University of Aix-Marseille II, Marseille, France. He then got the "Doctorat d'Etat es Sciences" degree in 1989 from the same University for his pioneer work on analysis and synthesis of sounds using time-frequency and time-scale (wavelets) representations. He is currently Director of Research at the National Center for Scientific Research (CNRS), Laboratoire de Mécanique et d'Acoustique, Marseille. His primary research interests are in analysis and synthesis of sounds with a particular emphasis on high-level control of synthesis processes. He recently addressed applications linked to musical interpretation and semantic description of sounds using a pluridisciplinary approach associating signal processing, physics, perception and cognition.