



Time-Frequency Fading Algorithms Based on Gabor Multipliers

Ama Marina Kreme, Valentin Emiya, Caroline Chaux, Bruno Torr sani

► To cite this version:

Ama Marina Kreme, Valentin Emiya, Caroline Chaux, Bruno Torr sani. Time-Frequency Fading Algorithms Based on Gabor Multipliers. IEEE Journal of Selected Topics in Signal Processing, 2021, 15 (1), pp.65-77. 10.1109/JSTSP.2020.3045938 . hal-02861427v2

HAL Id: hal-02861427

<https://hal.science/hal-02861427v2>

Submitted on 10 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin e au d p t et   la diffusion de documents scientifiques de niveau recherche, publi s ou non,  manant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv s.

Time-frequency fading algorithms based on Gabor multipliers

A. Marina Krémé, Valentin Emiya, Caroline Chaux, *Senior member, IEEE*, and Bruno Torr sani

Abstract—In this paper, we address a particular instance of time-frequency filter design, which we call Time-Frequency Fading (TFF). In TFF the only available information concerns the time-frequency localization of the component to be filtered out or attenuated: the signal of interest is supposed to be spread out in the time-frequency plane, whereas the perturbation signal is concentrated within a specified time-frequency region Ω .

The problem is formulated as an optimization problem designed to fade out the perturbation with accurate control on the fading level. The corresponding objective function involves a data fidelity term that aims to match the TF coefficients of the estimated signal to those of the observed signal outside the perturbation support. It also involves a penalty term that controls the energy of the reconstructed signal, within that region.

We obtain the closed-form solution of the problem which involves Gabor multipliers, i.e. linear operators of the pointwise product by a time-frequency transfer function called a mask. We study the TF localization properties of dominant eigenvectors of these Gabor multipliers, with particular attention to the case where the region Ω is a disjoint union of several sub-regions. The decay properties of eigenvalues naturally lead to reduced-rank approximations, and further approximations are obtained in the multiply connected region case. Also, we exploit random projection methods to speed up eigenvalue decompositions and rank reduction. This is implemented in two TFF algorithms, that cover the cases of single or multiple regions.

The efficiency of the proposed approach is demonstrated on several audio signals where the perturbations are filtered while leading to a good quality of signal reconstruction.

I. INTRODUCTION

WE consider the problem of restoration of a signal that has been degraded by an additive perturbation, in the special case where the latter is localized in specified regions of a time-frequency domain. A prototypical application is time-frequency editing, which may be described as follows: given a graphical TF representation, remove (graphically with a rubber) some TF components and reconstruct the signal. To our knowledge, this problem has not been really considered from the methodological point of view, even though it has been addressed in several application-driven problems.

In general the problem is solved empirically by *time-frequency masking* [1], [2], namely pointwise multiplication with a real-valued function called *mask*, binary or more sophisticated (Wiener-type filters). Time-frequency masking

may be seen as a particular instance of linear time-varying (LTV) filtering (see [3], [4], [5], [6] and references therein), where the filter design is done directly in a time-frequency domain. The design itself depends on the information available on the perturbation to be filtered out, that may consist of the single information on its time-frequency localization as in the problem at hand here or maybe more quantitative. Whatever the design, the resulting masked time-frequency transforms generally do not satisfy the consistency conditions associated with the transform (not every image corresponds to a time-frequency transform), unless consistency is imposed as a constraint as in [7].

In this work, we formulate the problem in a different way and propose a variational approach to design such time-frequency filters. We limit ourselves to the case of discrete Gabor transform [8], [9] (a.k.a. subsampled STFT), but the approach can easily be extended to other linear time-frequency or time-scale transforms from which perfect reconstruction is possible (such as wavelet transform, constant Q transform, S transform or others). The time-frequency filters are obtained by minimizing a suitable objective function, that involves binary masking, but the resulting filter is not time-frequency masking. We call such an approach *time-frequency fading*, as it makes it possible to fade out a given region in the time-frequency domain by an amount that can be tuned by a (set of) control parameter(s), in an analogy with classical audio fading (gradual change in loudness) or video (gradual change in color). This control parameter is actually a regularization parameter in the objective function. We also propose a data driven strategy for tuning this parameter. At this point, let us stress that the fading problem does not involve the estimation of the region to be filtered out, which is fixed. In this respect, fading is closer to time-frequency filter design than to single channel blind source separation, which often involves the construction of a mask, be it binary or real.

Our approach can be summarized as follows. Let \mathcal{T} denote a time-frequency transform, mapping any signal $x \in \mathbb{C}^L$ to $\mathcal{T}x \in l^2(\Lambda)$. Let us denote by $x^o \in \mathbb{R}^L$ the target signal and $y^o \in \mathbb{R}^L$ the signal to be faded out. y^o is assumed to be essentially localized in a specific region Ω (that may be a disjoint union of connected sub-regions Ω_p far away from each other) in the time-frequency domain Λ . We assume that sub-regions Ω_p (and thus Ω) are symmetric with respect to the null-frequency axis since signals are real-valued. We denote by $\bar{\Omega} = \Lambda \setminus \Omega$ the complement region. The observed signal z is thus given by $z = x^o + y^o$. The objective is, given Ω , to retrieve x^o from z or equivalently, to remove y^o from z . We

A. Marina Kr  m  , Caroline Chaux and Bruno Torr  sani are with Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France (Email: ama-marina.kreme@univ-amu.fr, caroline.chaux@univ-amu.fr, bruno.torresani@univ-amu.fr).

A. Marina Kr  m   and Valentin Emiya are with Aix Marseille Univ, Universit   de Toulon, CNRS, LIS, Marseille, France (Email: ama-marina.kreme@univ-amu.fr, valentin.emiya@lis-lab.fr).

A. Marina Kr  m   was supported by a PhD grant from R  gion Sud.

formulate the problem as the following quadratic optimization problem

$$\mathbf{x}_\lambda = \underset{\mathbf{x} \in \mathbb{R}^L}{\operatorname{argmin}} \quad \|\mathcal{T}\mathbf{x} - \mathcal{T}\mathbf{z}\|_\Omega^2 + \sum_{p=1}^P \lambda_p \|\mathcal{T}\mathbf{x}\|_{\Omega_p}^2. \quad (1)$$

where $\|\mathbf{x}\|_\Omega^2 := \sum_{k \in \Omega} |\mathbf{x}[k]|^2$ and $\boldsymbol{\lambda} = (\lambda_p)_p$, $\lambda_p > 0$. This amounts to find an estimation \mathbf{x}_λ of \mathbf{x}^o which best matches \mathbf{z} outside Ω and whose energy within Ω is reduced. The problem (1) leads to a linear system, whose solution involves time-frequency multipliers (i.e. time-frequency filters defined by pointwise multiplication in the time-frequency domain), as shown in [10], where a spectral approach based upon eigenvalue decomposition of Gabor multipliers was proposed. However, this approach turns out to be computationally expensive with large signals. Furthermore, the approach was restricted to the case $P = 1$.

In this work, we build on the results of [10] and extend the approach to the case $P > 1$, still using Gabor multipliers. We propose several strategies to reduce the computational cost and thus enable real audio signal processing. We also propose a complete pipeline involving the splitting of Ω into pairwise disjoint sub-regions Ω_p and the estimation of optimal values for hyperparameters λ_p .

The paper is organized as follow: Section II is devoted to the introduction of Gabor tools while Section III presents the proposed method. The problems of erasing one region or several sub-regions are treated as well as efficient algorithmic issues. Numerical experiments are drawn in Section IV on various audio signals for which the proposed algorithm outperforms baseline approaches. Finally, Section V concludes the paper. More technical aspects are postponed to the appendix.

II. ELEMENTS OF GABOR ANALYSIS

Gabor analysis is a particular instance of time-frequency analysis, that aims at representing signals (functions, sequences, or vectors) as linear combinations of translated and modulated copies of a generic signal called a window. The coefficients of the expansion provide an alternative representation of the object of interest, called a time-frequency representation. Gabor analysis has enjoyed important developments since its introduction in the seminal paper by D. Gabor [8]. We refer to [11] for a detailed account of mathematical developments, which also addresses extensions to operator theory that will be of interest here, and to [12], [13], [14] for signal processing related developments. Gabor analysis has found many applications in various areas of signal processing such as audio, speech and image processing, communication theory, radar and sonar [15], [8]. We give below a short account of the main aspects that are important for this paper.

Let us first introduce some notations. Throughout this paper, matrices and vectors (considered columnwise) are represented using bold characters type (such as \mathbf{X} or \mathbf{x}), and scalars are represented using normal characters; the i th entry of \mathbf{x} is $\mathbf{x}[i]$, similar notations are used for matrix elements. The inner product of two column vectors \mathbf{x} and \mathbf{y} is defined by $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^H \mathbf{x}$, where H denotes Hermitian conjugation, and the standard norm is denoted by $\|\cdot\|_2$.

Given a matrix \mathbf{X} , we denote by $\|\mathbf{X}\|_2 = \sqrt{\operatorname{Tr}(\mathbf{X}^H \mathbf{X})}$ its Frobenius norm (with Tr denoting the trace) and by $\|\mathbf{X}\|$ its operator norm. Given a subset Ω of the matrix index set, the Frobenius norm of the restriction of \mathbf{X} to Ω is denoted by $\|\mathbf{X}\|_\Omega = \|\mathbf{X} \mathbf{1}_\Omega\|_2$, $\mathbf{1}_\Omega$ being the indicator function of Ω .

A. Gabor analysis

We limit ourselves to finite dimensional situations and describe Gabor analysis on \mathbb{C}^L , where L is a positive integer. Let a and b be two divisors of L , we set $N = L/a$ and $M = L/b$ and introduce the time-frequency lattice $\Lambda^\circ = b\mathbb{Z}_M \times a\mathbb{Z}_N$ and the lattice $\Lambda = \mathbb{Z}_M \times \mathbb{Z}_N$ (which we will use as index set for Gabor analysis), where $\mathbb{Z}_N := \mathbb{Z}/N\mathbb{Z} = \{0, 1, \dots, N-1\}$ is the cyclic group with N elements. Let $\mathbf{g} \in \mathbb{R}^L$ be a nonzero vector, hereafter called the *analysis window*. The family of Gabor atoms associated with $(\mathbf{g}, \Lambda^\circ)$ is the family of vectors $\mathbf{g}_{mn} \in \mathbb{C}^L$ obtained by translations and modulations of \mathbf{g} on the lattice Λ° . Namely, the Gabor atoms \mathbf{g}_{mn} are defined by

$$\mathbf{g}_{mn}[l] = \mathbf{g}[l - na]e^{2i\pi mbl/L}, \quad l \in \mathbb{Z}_L$$

A family of Gabor atoms as above is a *Gabor frame* when there exists two constants $0 < A \leq B < \infty$ verifying, for all $\mathbf{x} \in \mathbb{C}^L$,

$$A\|\mathbf{x}\|_2^2 \leq \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} |\langle \mathbf{x}, \mathbf{g}_{mn} \rangle|^2 \leq B\|\mathbf{x}\|_2^2 \quad (2)$$

The constants A and B are respectively the lower and upper frame bounds. When $A = B$, the Gabor frame is *tight*, and is called a Parseval frame when $A = B = 1$. The redundancy of the representation is $L/ab = MN/L$, it has to be greater than or equal to 1 to allow perfect reconstruction [11], [16]. However, the Balian-Low theorem states that if \mathbf{g} is to be sufficiently localized in time and frequency, which is the case in most applications including the present paper, the redundancy must be greater than 1.

The discrete Gabor transform (DGT) [17], [18] \mathcal{V} maps any $\mathbf{x} \in \mathbb{C}^L$ into a matrix $\mathcal{V}\mathbf{x} \in \mathbb{C}^{M \times N}$ of *Gabor coefficients*, defined by

$$\mathcal{V}\mathbf{x}[m, n] = \langle \mathbf{x}, \mathbf{g}_{mn} \rangle = \sum_{l=0}^{L-1} \mathbf{x}[l] \mathbf{g}[l - na] e^{-2i\pi mbl/L}.$$

The spectrogram is the squared modulus $|\mathcal{V}\mathbf{x}[m, n]|^2$ of the DGT, often represented in logarithmic scale.

When the signal \mathbf{x} is real-valued, its DGT is Hermitian, i.e., $\mathcal{V}\mathbf{x}[M-m, n] = \overline{\mathcal{V}\mathbf{x}[m, n]}$, $\forall m \in \{1, \dots, M-1\}$. In such a case it is customary to represent non-negative frequencies only.

The adjoint operator $\mathcal{V}^H : \mathbb{C}^{M \times N} \rightarrow \mathbb{C}^L$, also called synthesis operator is given as follows. For all $\mathbf{c} \in \mathbb{C}^{M \times N}$,

$$\mathcal{V}^H \mathbf{c} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \mathbf{c}[m, n] \mathbf{g}_{mn}.$$

The frame operator is defined by $\mathbf{S} = \mathcal{V}^H \mathcal{V}$. It is bounded, self-adjoint and semi positive definite (i.e. $\mathbf{x}^H \mathbf{S} \mathbf{x} \geq 0$ for $\mathbf{x} \in \mathbb{C}^L$) by construction, and actually positive definite as a

consequence of (2). Therefore \mathbf{S} is invertible, which permits to reconstruct any $\mathbf{x} \in \mathbb{C}^L$ from its Gabor coefficients.

Of particular interest are the above mentioned Parseval frames, for which $\mathbf{S} = \mathbf{I}$, i.e., \mathcal{V}^H is a left inverse of \mathcal{V} . In such situations, the Gabor expansion takes a very simple form: for all $\mathbf{x} \in \mathbb{C}^L$

$$\mathbf{x} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \langle \mathbf{x}, \mathbf{g}_{mn} \rangle \mathbf{g}_{mn}.$$

For the sake of simplicity, we will limit ourselves to this case throughout this paper.

Parseval Gabor frames can easily be generated from any Gabor frame. Indeed, it can be shown that given a Gabor frame associated with $(\mathbf{g}, \Lambda^\circ)$, with frame operator \mathbf{S} , and defining a new window $\gamma = \mathbf{S}^{-1/2} \mathbf{g}$, the Gabor frame associated with (γ, Λ°) is a Parseval Gabor frame. In this paper we will focus on Parseval frames constructed in this way, starting from Gaussian and Hann windows.

B. Gabor multipliers

Gabor multipliers are linear operators that perform time-varying signal filtering by pointwise multiplication in the Gabor domain.

Definition II.1 (Gabor multiplier). *Let \mathbf{g} and Λ° be as above, and let $\mathbf{m} \in \mathbb{C}^{M \times N}$. The Gabor multiplier associated to $(\mathbf{g}, \Lambda^\circ)$ and \mathbf{m} , is the linear operator $\mathfrak{M}_{\mathbf{m}} = \mathcal{V}^H \mathbf{m} \mathcal{V}$, acting on $\mathbf{x} \in \mathbb{C}^L$ as*

$$\mathfrak{M}_{\mathbf{m}} \mathbf{x} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \mathbf{m}[m, n] \langle \mathbf{x}, \mathbf{g}_{mn} \rangle \mathbf{g}_{mn}.$$

\mathbf{m} is called the mask (or time-frequency transfer function) of the Gabor multiplier¹.

Gabor multipliers form a particular class of time-varying filters [19], [4], which is relevant in various domains including time-varying system identification [20] or modulation theory [21] or sound morphing [22]. Approximation of linear systems by Gabor multipliers have been studied in various places (see e.g. [23], [24], [25] and references therein). The main property is that they are good at approximating so-called underspread linear systems, i.e. linear systems that do not involve large time shifts and frequency shifts.

The class of Gabor multipliers considered in this paper enjoy useful properties, which we address below. We first start with generic properties.

Properties II.1 ([19], [26]). (i) *If \mathbf{m} is real-valued then $\mathfrak{M}_{\mathbf{m}}$ is self-adjoint. It is then diagonalizable, with real eigenvalues, and there exists an orthonormal basis of \mathbb{C}^L formed by $\mathfrak{M}_{\mathbf{m}}$ eigenvectors.* (ii) *The Gabor multiplier generated by $\mathbf{m} \equiv 1$ is a multiple of the identity operator if and only if $(\mathbf{g}, \Lambda^\circ)$ generates a tight Gabor frame.*

(iii) *For arbitrary $\mathbf{m} \in \mathbb{C}^{M \times N}$, then $\mathfrak{M}_{\mathbf{m}}$ defines a bounded operator with operator norm $\|\mathfrak{M}_{\mathbf{m}}\| \leq B \|\mathbf{m}\|_\infty$, where B is the upper bound defined in (2). In particular, if \mathbf{g} and Λ° generate a Parseval frame, then $\|\mathfrak{M}_{\mathbf{m}}\| \leq \|\mathbf{m}\|_\infty$.*

We will mainly focus on the so-called time-frequency localization operators, which are Gabor multipliers associated with a mask $\mathbf{m} = 1_\Omega$ equal to the indicator function of a region Ω of the time-frequency lattice Λ . The mask \mathbf{m} being binary, and since we limit the present investigations to Parseval frames, the corresponding eigenvalues range between 0 and 1.

Time-frequency localization properties of corresponding eigenvectors (which can be seen as generalized prolate spheroidal functions) have been studied in [27], [19], see [28] and references therein for recent accounts. The main result is that when the area of the region Ω is large enough, eigenvectors with large eigenvalues (i.e. close to 1) tend to be concentrated within Ω , while eigenvectors with smaller eigenvalues tend to localize themselves outside Ω . Therefore, the rank (or effective rank, i.e. dimension of the subspace generated by eigenvectors with significant eigenvalues) grows with the area of Ω (we refer to [28] for more precise statements). We illustrate these properties in the next subsection.

C. Numerical illustrations

We illustrate in this section the behavior of Gabor transform and Gabor multipliers in situations of interest here. We rely on dedicated toolboxes, namely the LTFAT Octave/Matlab toolbox [29], [30], and the Python version ltfatpy [31].

We first highlight the behavior of eigenvectors and eigenvalues of Gabor multipliers on a mask generated from the sum of a car engine sound and a bird song with length 16384 (see Section IV for the full experimental setting). The mask was constructed by appropriate thresholding of the Gabor transform of the resulting signal, followed by simple morphological operations that ensure sufficient connectivity of sub-regions (see Section IV-A below for details). The mask is displayed in Fig 1 (left panel), the eigenvalues of the corresponding Gabor multiplier are displayed on the right panel, sorted in decreasing order. Two windows have been considered, respectively generated from a Gaussian and Hann window (we recall that initial windows are modified so as to generate Parseval frames).

As expected, eigenvalues range between 0 and 1. A plateau of values equal to 1 can be observed at the beginning, followed by a rapid decay. As an example, less than 3900 eigenvalues (out of 16384) are enough to reach a precision of 10^{-6} for the Gaussian-type window (3700 for Hann), which opens the door to significant dimension reduction. We also illustrate in Fig. 2 the time-frequency localization of eigenvectors related to large and smaller eigenvalues. It can be seen on the top row that eigenvectors associated with the top eigenvalue $\sigma[1]$ (equal to 1 with excellent precision, namely 2.4×10^{-11} for the Gaussian type window, and 3.4×10^{-12} for the Hann type window)

tend to localize in the same regions as one of the connected sub-regions of the mask (here sub-regions 1 and 8, from left to right). Eigenvectors associated with lower eigenvalues are

¹By abuse of notation we have used the same symbol to represent the mask and the operator of pointwise multiplication by the mask.

displayed on the bottom row, (chosen so as to have close eigenvalues, and approximately the same TF localization as the top eigenvector: $\sigma[3072] \approx 3.914 \times 10^{-4}$ and $\sigma[3199] \approx 3.865 \times 10^{-4}$) tend to be localized within time-frequency “rings” around the connected components of the mask, the smaller the eigenvalue the larger the ring radius.

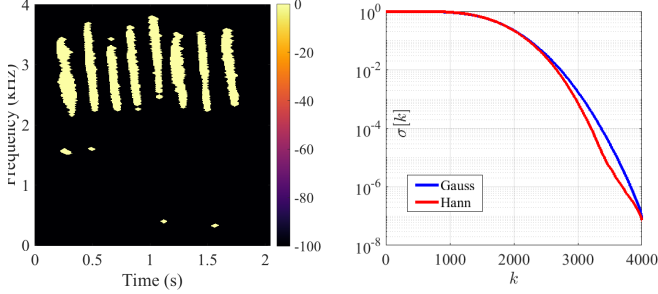


Fig. 1. Binary Mask generated with a Gauss window estimated from a bird song signal (left), and Gabor multiplier’s eigenvalues $\sigma[k]$ in decreasing order, displayed in logarithmic scale (right).

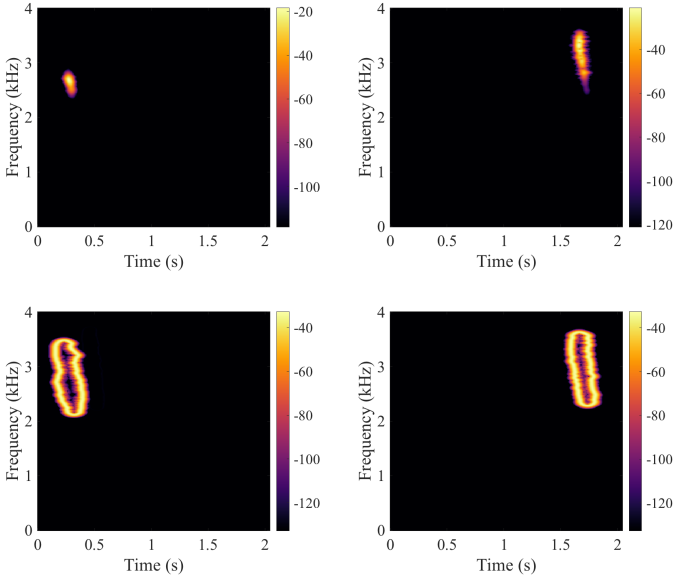


Fig. 2. Column 1 (top to bottom): spectrogram of the first and 3072 th eigenvectors of the Gabor multiplier generated with a Gauss window. Column 2 (top to bottom): spectrogram of the first and 3199 th eigenvectors of the Gabor multiplier generated with a Hann window. The mask associated with each of these multipliers was generated using the Gaussian type window, as shown in Fig. 1.

The relationship between the support size of the mask and the behavior of eigenvalues of the corresponding Gabor multiplier has been studied theoretically by several authors, in a continuous time setting (see e.g. [28] and references therein). Assuming one may extrapolate to the discrete, finite-dimensional case, one would expect the number of eigenvalues larger than some threshold to vary (at least approximately) linearly with the support size. We checked this property numerically in situations such as the ones we consider in this paper.

More precisely, we computed a rank estimate for Gabor multipliers with a binary mask with rectangular support, constant shape and center, and increasing areas. Namely, we computed the number of eigenvalues above a fixed threshold (set here to 10^{-13}), as a function of the area of the support of the mask. We use a signal length $L = 16384$, a Gaussian window with length 64, a hop size $a = 16$ and $b = 256$ bins. Results are displayed in Fig. 3. Within the range of interest, the rank estimate appears to vary linearly as a function of the mask area. Fig. 3 also includes another estimate called Rand-EVD and based upon random projections, which we use and describe in Section III-C2 below. This estimate exhibits the same linear behavior; we remark that Rand-EVD tends to slightly underestimate the rank compared to the reference approach.

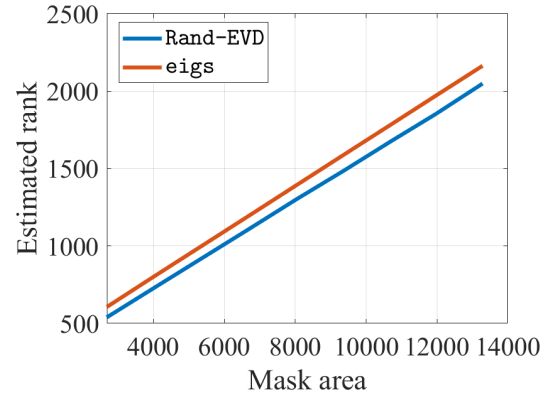


Fig. 3. Rank estimation as a function of mask area. The window is a *Gaussian* window.

III. TIME-FREQUENCY FADING (TFF)

A. Problem statement

As mentioned in Section I, we assume that we observe a signal z of length L , of the form

$$z = x^o + y^o, \quad (3)$$

i.e., the sum of a signal of interest x^o and a perturbation y^o . We assume further that the perturbation is strongly concentrated within a known region $\Omega \subset \Lambda$ in the time-frequency (TF) domain Λ . We recall that $\bar{\Omega} = \Lambda \setminus \Omega$ is the complement region of Ω .

Our objective is to filter out the perturbation y^o within Ω and reduce the artifacts generated by filtering. We term this problem *time-frequency fading* (TFF).

The example in Figure 1 suggests to split the region Ω into smaller disjoint sub-regions Ω_p , $1 \leq p \leq P$ and attempt to approximate the problem by P independent subproblems. As we will see, this turns out to be possible if sub-regions Ω_p are sufficiently far away from each other, with respect to a distance measure to be properly defined. We provide in Section III-B3 below and the Appendix a more quantitative analysis that also justifies one of the algorithms we propose (TFF-P).

Standard practice for such a TFF problem is to perform a pointwise multiplication in the time-frequency domain by

some transfer function (see e.g. [1], [32], [16], [33], [34] for applications in various domains), that penalizes the component within Ω , i.e. using a Gabor multiplier. However the choice of the transfer function is often arbitrary, and standard choices (such as binary masks) often generate artifacts, as stressed in [7].

We choose here to formulate the problem as an optimization problem, which we write as

$$\mathbf{x}_\lambda = \underset{\mathbf{x} \in \mathbb{R}^L}{\operatorname{argmin}} f_\lambda(\mathbf{x}) := \|\mathcal{V}\mathbf{x} - \mathcal{V}\mathbf{z}\|_\Omega^2 + \sum_{p=1}^P \lambda_p \|\mathcal{V}\mathbf{x}\|_{\Omega_p}^2 \quad (4)$$

where the objective function f_λ depends on regularization parameters $\lambda_p > 0$ for all sub-regions $p = 1, \dots, P$, which we collect in a vector $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_P)$.

The first term of the objective function in (4) is a data fidelity term that matches the DGT of the estimated signal to that of the observation outside Ω . The second term controls its energy in each sub-region Ω_p , and the regularization parameters control the trade-off among all terms. The importance of the hyperparameters is discussed in Remark III.2 below.

The objective function in (4) is a quadratic form, its optimization leads to a linear system which we study below.

Remark III.1. In the case $P = 1$, the problem can also be formulated in terms of the perturbation \mathbf{y} in the observation model (3). Indeed, one may also write

$$g_\lambda(\mathbf{y}) = \frac{1}{\lambda} f_\lambda(\mathbf{z} - \mathbf{y}) = \|\mathcal{V}\mathbf{y} - \mathcal{V}\mathbf{z}\|_\Omega^2 + \frac{1}{\lambda} \|\mathcal{V}\mathbf{y}\|_\Omega^2$$

so that optimizing f_λ with respect to \mathbf{x} is equivalent to optimizing g_λ with respect to \mathbf{y} . Despite this apparent symmetry, the problems of reconstructing \mathbf{x} and reconstructing \mathbf{y} are not equivalent. Indeed, we showed in [10] that the optimal value of λ for reconstructing \mathbf{x} is not the same as the optimal value for reconstructing \mathbf{y} , mainly due to the imbalance of sizes of Ω and $\bar{\Omega}$. A strategy for tuning λ is described in Section III-C3 below.

B. Proposed solutions

1) *Closed-form solution:* Because f_λ is quadratic, solutions of the optimization problem (4) are given by the corresponding critical points. The gradient of f_λ can be expressed in terms of Gabor multipliers

$$\nabla f_\lambda(\mathbf{x}) = 2\mathfrak{M}_{\bar{\Omega}}\mathbf{x} - 2\mathfrak{M}_{\bar{\Omega}}\mathbf{z} + 2 \sum_{p=1}^P \lambda_p \mathfrak{M}_{\Omega_p} \mathbf{x}, \quad (5)$$

where $\mathfrak{M}_{\mathcal{D}}$ denotes the Gabor multiplier associated with the indicator function of the domain \mathcal{D} .

Since we limit our interest to the case of Parseval Gabor frames, we have $\mathfrak{M}_{\Omega} + \mathfrak{M}_{\bar{\Omega}} = \mathbf{I}$. Therefore, the normal equations can be written as

$$\left(\mathbf{I} - \sum_{p=1}^P (1 - \lambda_p) \mathfrak{M}_{\Omega_p} \right) \mathbf{x} = \left(\mathbf{I} - \sum_{p=1}^P \mathfrak{M}_{\Omega_p} \right) \mathbf{z}, \quad (6)$$

where we have also used the fact that the Ω_p are pairwise disjoint.

Remark III.2. Notice that setting $\lambda_p = 0$ in the above equation for every p leads to $\mathfrak{M}_{\bar{\Omega}}\mathbf{z} = \mathfrak{M}_{\bar{\Omega}}\mathbf{x}$, which does not have a unique solution if $\mathfrak{M}_{\bar{\Omega}}$ is not full rank; what happens within Ω is then uncontrolled. The same may be true if one of the constants λ_p is set to zero and if the corresponding sub-region Ω_p is large enough.

The operator $\mathbf{I} - \sum_{p=1}^P (1 - \lambda_p) \mathfrak{M}_{\Omega_p}$ is actually a Gabor multiplier with piecewise constant mask $\mathbf{m} = 1 - \sum_{p=1}^P (1 - \lambda_p) 1_{\Omega_p}$, where 1_{Ω_p} is the indicator function of the sub-region Ω_p . If $\lambda_p > 0$ for all p , then $\min_{(m,n)} (\mathbf{m}[m,n]) > 0$ and the mask is therefore semi-normalized, which implies invertibility (see Proposition 3.7 in [35]).

From now on we then assume $\lambda_p > 0$ for all p . The solution then reads

$$\begin{aligned} \mathbf{x}_\lambda &= \left(\mathbf{I} - \sum_{p=1}^P (1 - \lambda_p) \mathfrak{M}_{\Omega_p} \right)^{-1} \left(\mathbf{I} - \sum_{p=1}^P \mathfrak{M}_{\Omega_p} \right) \mathbf{z} \\ &= \mathbf{z} - \left(\mathbf{I} - \sum_{p=1}^P (1 - \lambda_p) \mathfrak{M}_{\Omega_p} \right)^{-1} \sum_{p=1}^P \lambda_p \mathfrak{M}_{\Omega_p} \mathbf{z}. \end{aligned} \quad (7)$$

This closed-form solution requires solving a linear system for each choice of $\boldsymbol{\lambda}$, which may be computationally demanding in high dimensions and/or when the region Ω is large. Indeed, owing to the discussion at the end of subsection II-A, the rank of the multipliers increases with the area of the support of the mask. We propose below efficient algorithmic solutions that provide good approximations of (7).

2) *Reduced rank approximation for $P = 1$:* Let us first examine the case where a single region Ω is considered (for simplicity we then drop the subregion index). Since $P = 1$, a unique regularization parameter $\lambda \in \mathbb{R}_+^*$ is required.

The solution reads

$$\mathbf{x}_\lambda = \mathbf{z} - \left(\mathbf{I} - (1 - \lambda) \mathfrak{M}_{\Omega} \right)^{-1} \lambda \mathfrak{M}_{\Omega} \mathbf{z}, \quad (8)$$

the invertibility of $\mathbf{I} - (1 - \lambda) \mathfrak{M}_{\Omega}$ being ensured by the fact that the spectrum of \mathfrak{M}_{Ω} is included in $[0, 1]$ (see below).

As mentioned in Properties II.1, the mask being binary, the Gabor multiplier \mathfrak{M}_{Ω} can be diagonalized and thus factored in matrix form as $\mathfrak{M}_{\Omega} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{U}^H$ where \mathbf{U} is unitary and $\boldsymbol{\Sigma} = \operatorname{diag}(\boldsymbol{\sigma})$ is diagonal. Here $\boldsymbol{\sigma} = (\sigma[1], \dots, \sigma[L]) \in \mathbb{R}^L$ is the vector of eigenvalues, which are sorted in decreasing order $1 \geq \sigma[1] \geq \dots \geq \sigma[L] \geq 0$.

We can then write

$$\mathbf{x}_\lambda = \mathbf{z} - \mathbf{U} \boldsymbol{\Gamma} \mathbf{U}^H \mathbf{z} \quad (9)$$

where $\boldsymbol{\Gamma} = \operatorname{diag}(\boldsymbol{\gamma})$, with $\boldsymbol{\gamma} = (\gamma[1], \dots, \gamma[L])$ and

$$\gamma[i] = \frac{\lambda \sigma[i]}{1 - (1 - \lambda) \sigma[i]}.$$

It is easily seen that the sequence $\boldsymbol{\gamma} = (\gamma[1], \dots, \gamma[L])$ is decreasing and included in $[0, 1]$, which we will use for truncation purposes.

The approximate solution when truncating to the K largest eigenvalues ($K < L$) is given by:

$$\mathbf{x}_\lambda^{(K)} = \mathbf{z} - \mathbf{U}^{(K)} \boldsymbol{\Gamma}^{(K)} (\mathbf{U}^{(K)})^H \mathbf{z} \quad (10)$$

where $\mathbf{U}^{(K)}$ is the $L \times K$ submatrix consisting in the first K columns of \mathbf{U} , and $\mathbf{\Gamma}^{(K)}$ is the submatrix consisting of the K first rows and columns of $\mathbf{\Gamma}$. The truncation error can be bounded as [10]

$$\|\mathbf{x}_\lambda - \mathbf{x}_\lambda^{(K)}\|_2 \leq \gamma[K+1] \|\mathbf{z}\|_2. \quad (11)$$

3) Approximate solution in the general case $P \geq 1$:

In the general case where P sub-regions are considered, we propose an approximation, given in eq. (13), of the practically-untractable closed-form solution (7). The approximation relies on the assumption that the composition of Gabor multipliers related to two different sub-regions may be neglected provided the regions are sufficiently far apart:

$$p \neq q \Rightarrow \|\mathfrak{M}_{\Omega_p} \mathfrak{M}_{\Omega_q}\| \approx 0, \forall p, q. \quad (12)$$

Let us assume for the sake of the argument² that for all $p \neq q$, $\|\mathfrak{M}_{\Omega_p} \mathfrak{M}_{\Omega_q}\| = 0$, in other words the range of \mathfrak{M}_{Ω_q} is in the null space of \mathfrak{M}_{Ω_p} ; as a result, the eigenvectors of \mathfrak{M}_{Ω_q} with the largest eigenvalues are orthogonal to the eigenvectors of \mathfrak{M}_{Ω_p} with the largest eigenvalues.

Hence one may jointly diagonalize all the Gabor multipliers $\{\mathfrak{M}_{\Omega_p}\}_{p=1}^P$ in a common orthonormal basis of eigenvectors $\mathbf{U}^{(K)} := [\mathbf{U}_1^{(K_1)}, \dots, \mathbf{U}_P^{(K_P)}, \mathbf{U}_0^{(K_0)}]$ where \mathbf{U}_p is the matrix of orthonormal eigenvectors associated with the K_p largest eigenvalues $\sigma_p^{(K_p)}$ of \mathfrak{M}_{Ω_p} sorted in decreasing order, and where $\mathbf{U}_0^{(K_0)}$ is composed of $K_0 = L - \sum_{p=1}^P K_p$ orthonormal vectors to complete the basis. This is valid when the K_p are small enough, i.e., when the perturbation is well localized so that regions Ω_p are small. It follows that the closed form solution (7) is approximated by

$$\begin{aligned} \mathbf{x}_\lambda^{(K)} &= \mathbf{z} - \left(\mathbf{I} - \mathbf{U}^{(K)} \text{diag} \begin{bmatrix} \lambda_1 \sigma_1^{(K_1)} \\ \vdots \\ \lambda_P \sigma_P^{(K_P)} \\ \mathbf{0}_{K_0} \end{bmatrix} (\mathbf{U}^{(K)})^H \right)^{-1} \\ &\quad \times \sum_{p=1}^P \lambda_p \mathbf{U}_p^{(K_p)} \text{diag} \sigma_p^{(K_p)} (\mathbf{U}_p^{(K_p)})^H \mathbf{z} \\ &= \mathbf{z} - \sum_{p=1}^P \mathbf{U}_p^{(K_p)} \text{diag} (1 - \lambda_p \sigma_p^{(K_p)})^{-1} (\mathbf{U}_p^{(K_p)})^H \\ &\quad \times \sum_{p=1}^P \lambda_p \mathbf{U}_p^{(K_p)} \text{diag} \sigma_p^{(K_p)} (\mathbf{U}_p^{(K_p)})^H \mathbf{z} \end{aligned}$$

Defining $\mathbf{\Gamma}_p^{(K_p)} = \text{diag}(\gamma_p[1], \dots, \gamma_p[K_p])$ with $\gamma_p[i] =$

$\frac{\lambda_p \sigma_p[i]}{1 - \lambda_p \sigma_p[i]}$, we then obtain

$$\begin{aligned} \mathbf{x}_\lambda^{(K)} &= \mathbf{z} - \sum_{p=1}^P \mathbf{U}_p^{(K_p)} \mathbf{\Gamma}_p^{(K_p)} (\mathbf{U}_p^{(K_p)})^H \mathbf{z} \\ &\quad - \sum_{p \neq q} \mathbf{U}_p^{(K_p)} \text{diag} (1 - \lambda_p \sigma_p^{(K_p)})^{-1} (\mathbf{U}_p^{(K_p)})^H \mathbf{U}_q^{(K_q)} \\ &\quad \text{diag} (\lambda_q \sigma_q^{(K_q)}) (\mathbf{U}_q^{(K_q)})^H \mathbf{z}. \end{aligned}$$

Since $(\mathbf{U}_p^{(K_p)})^H \mathbf{U}_q^{(K_q)} \approx 0$ the approximate solution reads

$$\mathbf{x}_\lambda^{(K)} = \mathbf{z} - \sum_{p=1}^P \mathbf{U}_p^{(K_p)} \mathbf{\Gamma}_p^{(K_p)} (\mathbf{U}_p^{(K_p)})^H \mathbf{z}, \quad (13)$$

which is actually a truncated version of a sum of independent single region solutions (9)

$$\mathbf{x}_\lambda^L = \mathbf{z} - \sum_{p=1}^P (\mathbf{I} - (1 - \lambda_p) \mathfrak{M}_{\Omega_p})^{-1} \lambda_p \mathfrak{M}_{\Omega_p} \mathbf{z}. \quad (14)$$

In fact the norms $\|\mathfrak{M}_{\Omega_p} \mathfrak{M}_{\Omega_q}\|$ are only approximately zero. We show in the Appendix that the error induced by the approximation can actually be controlled, assuming $0 < \lambda_p < 1$ for all p .

Set $\tilde{\lambda}_p = \min_{q \neq p} \lambda_q$ and $\Delta_p = \sum_{q \neq p} \|\mathfrak{M}_{\Omega_p} \mathfrak{M}_{\Omega_q}\|$, and assume that for $1 \leq p \leq P$, $\Delta_p < \frac{\lambda_p}{1 - \lambda_p} \frac{\tilde{\lambda}_p}{1 - \tilde{\lambda}_p}$, i.e., that the products of pairs of different Gabor multipliers are small enough compared to the values of regularization parameters. Given the localization properties of eigenvectors, this typically happens when sub-regions Ω_p are sufficiently far away from each others, and/or when regularization parameters are close to 1. We then obtain the upper bound (see Appendix A for a proof).

$$\|\mathbf{x}_\lambda - \mathbf{x}_\lambda^L\|_2 \leq \sum_{p=1}^P \frac{(1 - \tilde{\lambda}_p) \Delta_p}{\lambda_p \tilde{\lambda}_p - (1 - \lambda_p)(1 - \tilde{\lambda}_p) \Delta_p} \|\mathbf{z}\|_2, \quad (15)$$

The truncation error estimate (11) can be extended to this new setting and reads

$$\|\mathbf{x}_\lambda^L - \mathbf{x}_\lambda^{(K)}\|_2 \leq \sum_{p=1}^P \gamma_p[K_p + 1] \|\mathbf{z}\|_2.$$

The triangle inequality then yields the bound

$$\|\mathbf{x}_\lambda - \mathbf{x}_\lambda^{(K)}\|_2 \leq C(\lambda, K) \|\mathbf{z}\|_2, \quad (16)$$

with

$$C(\lambda, K) = \sum_{p=1}^P \left(\frac{(1 - \tilde{\lambda}_p) \Delta_p}{\lambda_p \tilde{\lambda}_p - (1 - \lambda_p)(1 - \tilde{\lambda}_p) \Delta_p} + \gamma_p[K_p + 1] \right). \quad (17)$$

²In practice, this assumption never holds exactly (except for specific cases such as regions that are far away in time, using finite-length windows), so the proposed argument is an intuitive justification of our method rather than a formal proof. Yet, very low values of $\|\mathfrak{M}_{\Omega_p} \mathfrak{M}_{\Omega_q}\|$ are reached in practice (e.g. see the $\epsilon = 10^{-5}$ threshold used in the experiments).

C. All-in-one algorithms

Relying on solutions presented in Sections III-B2 and III-B3, we introduce Algorithms 1 and 2 that process an observation z and a mask Ω in order to estimate the target signal. Algorithm 1 (resp. 2), named TFF-1 (resp. TFF-P), implements the solution of Section III-B2 (resp. III-B3), by considering Ω as a whole region (resp. as multiple subregions). In particular, it includes the creation of the Gabor multipliers (steps 2 in Alg. 1 and 4 in Alg. 2), the computation of their approximate EVD (steps 3 in Alg. 1 and 5 in Alg. 2), and the construction of the solutions (steps 6 in Alg. 1 and 9 in Alg. 2).

In order to make them computationally efficient and convenient for real-world usage, those algorithms also rely on three additional aspects :

- in TFF-P, the TF region Ω is partitioned into subregions (step 1 in Alg. 2); this can be done automatically without specifying the number of subregions, as described in Section III-C1;
- the Gabor multipliers are diagonalized using fast randomized algorithms, as described in Section III-C2
- efficient hyperparameter tuning is obtained by estimating target energy values in the (sub-)regions (steps 1-4-5 in Alg. 1 and 2-6-7 in Alg. 2), as detailed in Section III-C3.

Algorithm 1 TFF-1: filtering out one TF region

Require: Mix z , mask Ω , DGT \mathcal{V} , parameters $\epsilon_{\text{EVD}}, p_{\text{EVD}}$,

- 1: Estimate target energy $e_{\text{target}} \in \mathbb{R}$ in Ω (see III-C3)
- 2: $\mathfrak{M}_{\Omega} \leftarrow \mathcal{V}^H 1_{\Omega} \mathcal{V}$
- 3: $\hat{U}, \hat{\sigma} \leftarrow \text{rand_evd}(\mathfrak{M}_{\Omega}, \epsilon_{\text{EVD}}, p_{\text{EVD}})$
- 4: Let

$$\hat{x} : \lambda \mapsto z - \hat{U} \Gamma(\lambda) \hat{U}^H z$$

where

$$\Gamma : \lambda \mapsto \text{diag} \left(\frac{\lambda \hat{\sigma}}{1 - (1 - \lambda) \hat{\sigma}} \right)$$

- 5: $\hat{\lambda} \leftarrow \argmin_{\lambda > 0} |e_{\text{target}} - \|\mathcal{V} \hat{x}(\lambda)\|_{\Omega}^2|$
 - 6: **return** $\hat{x}(\hat{\lambda})$
-

1) *Partitioning mask Ω into sub-regions:* As introduced in Section III-B3, the solution using P sub-regions $\{\Omega_p\}_{p=1}^P$ relies on the hypothesis that the norm $\|\mathfrak{M}_{\Omega_p} \mathfrak{M}_{\Omega_q}\|$ of the composition of Gabor multipliers from sub-regions Ω_p and Ω_q is small, which occurs when sub-regions are sufficiently far away from each other in the time-frequency plane, in the sense that Ω_p and Ω_q satisfy the condition $\|\mathfrak{M}_{\Omega_p} \mathfrak{M}_{\Omega_q}\| \leq \epsilon$, for some small ϵ . We propose in Algorithm 3 an approach that splits an initial region Ω into P such sub-regions. The starting point (step 1) is a partition of Ω into connected components, e.g. by clustering TF points that are immediate neighbors using a 4-connected structuring element as done in image processing³. Then, pairs of sub-regions (Ω_p, Ω_q) for which

Algorithm 2 TFF-P: filtering out P TF sub-regions

Require: Mix z , mask Ω , DGT \mathcal{V} , parameters $\epsilon_{\text{EVD}}, p_{\text{EVD}}$

- 1: Split Ω into sub-regions $\{\Omega_p\}_{p=1}^P$ (e.g. via Algorithm 3)
- 2: Estimate target energies $e_{\text{target}} \in \mathbb{R}^P$ in $\{\Omega_p\}_{p=1}^P$ (see III-C3)
- 3: **for** $p = 1$ to P **do**
- 4: $\mathfrak{M}_{\Omega_p} \leftarrow \mathcal{V}^H 1_{\Omega_p} \mathcal{V}$
- 5: $\hat{U}_p, \hat{\sigma}_p \leftarrow \text{rand_evd}(\mathfrak{M}_{\Omega_p}, \epsilon_{\text{EVD}}, p_{\text{EVD}})$
- 6: Let

$$\hat{x}_p : \lambda \mapsto z - \hat{U}_p \Gamma_p(\lambda) \hat{U}_p^H z$$

where

$$\Gamma_p : \lambda \mapsto \text{diag} \left(\frac{\lambda \hat{\sigma}_p}{1 - (1 - \lambda) \hat{\sigma}_p} \right)$$

- 7: $\lambda_p \leftarrow \argmin_{\lambda > 0} |e_{\text{target}}[p] - \|\mathcal{V} \hat{x}_p(\lambda)\|_{\Omega_p}^2|$
 - 8: **end for**
 - 9: **return** $z - \sum_{p=1}^P \hat{U}_p \Gamma_p(\lambda_p) \hat{U}_p^H z$
-

$\|\mathfrak{M}_{\Omega_p} \mathfrak{M}_{\Omega_q}\|$ is greater than a tolerance threshold are merged repeatedly until no such pairs of regions exist (step 3).

Algorithm 3 Finding sub-regions for TFF-P

Require: Mask Ω , tolerance $\epsilon > 0$ on the spectral norm of the composition of two Gabor multipliers, DGT \mathcal{V}

- 1: Generate a partition $\mathcal{P} = \{\Omega_p\}$ of Ω into disjoint connected components.
 - 2: **while** $\exists \Omega_p, \Omega_q \in \mathcal{P}, \|\mathfrak{M}_{\Omega_p} \mathfrak{M}_{\Omega_q}\| > \epsilon$ **do**
 - 3: For every p, q such that $\|\mathfrak{M}_{\Omega_p} \mathfrak{M}_{\Omega_q}\| > \epsilon$, replace Ω_p and Ω_q by $\Omega_p \cup \Omega_q$ in \mathcal{P}
 - 4: **end while**
 - 5: **return** \mathcal{P}
-

For the bird+car example mentioned in Fig. 1, running Algorithm 3 with $\epsilon = 10^{-5}$ yields 9 sub-regions for which the actual values of $\|\mathfrak{M}_{\Omega_p} \mathfrak{M}_{\Omega_q}\|$ are of the order of 10^{-29} . This demonstrates the practical suitability of the sub-regions splitting.

2) *Diagonalization of Gabor multipliers using Rand-EVD:* The algorithms proposed in this paper rely strongly on eigenvalue decompositions (EVD) of Gabor multipliers. However, most EVD algorithms have complexity $\mathcal{O}(L^3)$, which becomes prohibitive in high dimensions. This can be reduced when the numerical rank K is significantly lower than the dimension. This is the case here, at least when the region Ω is significantly smaller than the time-frequency domain Λ , as shown in Section II-C. Provided that K is known, such a low-rank approximation is computed in $\mathcal{O}(KL^2)$ by well-known algorithms based on the Lanczos algorithm such as the Implicitly Restarted Arnoldi Method available in standard libraries under the usual name `eigs` [36], [37]. Even more efficient algorithms exist, as described as in [38], based on random projections. In addition to being significantly faster, they are also able to estimate the numerical rank K while the previous methods need to be given this rank. In

³See `bwlabel` in Matlab or `scipy.ndimage.label` in Python.

the current context, the low-rank eigenvalue decomposition of Gabor multipliers is obtained by combining the Adaptive Range Finder Algorithm [38, Algorithm 4.2] followed by the so-called Eigenvalue Decomposition via Nystrom Method [38, Algorithm 5.5]. The Adaptive Range Finder Algorithm provides a low-rank matrix that approximates the range of the Gabor multiplier. The rank is estimated from a target approximation error ϵ_{EVD} and a probability p_{EVD} to reach this error. This algorithm is particularly suitable in the case of fast operators with fast-decreasing eigenvalues like the DGT-based multipliers used here. Then, the Eigenvalue Decomposition via Nystrom Method uses this matrix to compute the approximate EVD, taking advantage of the Hermitian, positive semidefinite properties of our Gabor multipliers.

The combination of both algorithms will be named Rand-EVD in the rest of the paper. Rand-EVD [38] allows us to overcome precision and efficiency problems when performing classical EVD calculations in large dimensions. In practice, we have observed that Rand-EVD is able to estimate the numerical rank and to compute an approximate EVD about three times faster than `eigs` in settings such as those described in Section IV with parameters $\epsilon_{\text{EVD}} < 10^{-3}$ and $p_{\text{EVD}} = 0.9999$.

3) *Efficient solutions for tuning hyperparameters:* We propose a simple method to estimate hyperparameter λ . The main idea is to consider a target energy for the coefficients within Ω or any sub-region Ω_p , and to adjust the related hyperparameter consequently to fit this target value. Indeed, as can be seen in Problem (1), the higher the value of λ_p , the lower the energy $\|\mathcal{V}\mathbf{x}\|_{\Omega_p}^2$ in sub-region Ω_p .

The target energy may be adjusted by hand or by computing the energy in an unmasked region similar to the masked one, as described in [10]. Here, we propose a more systematic and yet simple strategy, assuming that the target signal has some stationarity properties: for each frequency channel f , we compute the average energy $\bar{e}_f = \sum_{t:(t,f) \in \Omega} |\mathcal{V}\mathbf{z}[t, f]|^2$ of all the unmasked coefficients $\mathcal{V}\mathbf{z}$ in frequency bin f ; then, we estimate the energy in Ω_p as $e_{\text{target}}[p] = \sum_{(t,f) \in \Omega_p} \bar{e}_f$; the energy in Ω is estimated as $e_{\text{target}} = \sum_{p=1}^P e_{\text{target}}[p]$.

Given target energies computed by the proposed strategy or by any other method, one can tune the hyperparameters by solving the optimization problems at step 5 in Algorithm 1 or at step 7 in Algorithm 2. This is computationally efficient for two reasons. First, while usual hyperparameter tuning via cross validation [39] generally implies to run the whole optimization procedure for model estimation at each hyperparameter evaluation, TFF-1 and TFF-P requires to diagonalize the Gabor multipliers only once. Hence, the cost of computing a solution for a given λ or λ_p is linear in the signal length. Second, estimating each λ_p is done separately thanks to the fact that sub-regions are far away from each other by construction (see Section III-C1). Hence, simple line searches over each λ_p in λ are conducted, avoiding a costly global P-dimensional optimization over λ .

An illustration in the case of a bird+car signal previously mentioned is given in Section IV-B and in Figure 5.

IV. NUMERICAL EXPERIMENTS

This Section is focused on practical experiments with the algorithms introduced in Section III. In Section IV-A, the experimental setting is introduced, including the sound material, the problem generation, the performance measure, the solvers, their parameters, and implementation details. In Section IV-B, we show and comment on results for the time-frequency fading problem using five different approaches (2 baseline one and 3 based on the proposed approaches):

For reproducibility, all the code and data are available in the `tff2020` toolkit⁴. Simulations have been conducted both in MATLAB and Python on CPU at 2.3GHz.

A. Experimental setting

1) *Sound material:* all audio signals considered here have been sampled at 8 kHz and their length is $L = 16384$ samples (i.e., about 2 seconds). The bird and car signals were provided by our partner ANSYS [40], chirps and modulations signals were synthetically created in our toolboxes. All the other signals have been downloaded from the Freesound database⁵.

The observed signal $\mathbf{z} = \mathbf{x}^\circ + \mathbf{y}^\circ$ is constructed as the sum of a target signal \mathbf{x}° and a perturbation signal \mathbf{y}° . The target signals have a wideband spectrogram while the perturbation signals are well localized in time-frequency. Original signals are normalized so that the target signal energy is 8dB above the energy in the perturbation signals after adding them up. We consider 18 such observed signals obtained from all pairs of target signal (*car*, *plane* and *train* engine sounds) and perturbation signal (*beeps*, *bird* song, *chirps*, *clicks*, *finger snaps* and *modulations*).

2) *Problem generation:* the DGT is parameterized using 1024 frequency bins and a hop size is $a = 64$. To favor the TF localization properties of the Gabor multiplier eigenvectors as highlighted in Section II, we used a Gaussian window for which 96% of the energy is concentrated within a width of 256 samples. Considering only non-negative frequencies, it results in a time-frequency matrix of size 513×256 . With these parameters, the redundancy of the Gabor frame equals $MN/L = 16$. Some additional experiments have also been conducted using a Hann window of size 512 samples. Results are similar to those with the Gaussian window and are not reported here in detail.

The TF fading problem is based upon the knowledge of the region Ω . In our experiments, Ω was generated from the spectrograms of original signals $S_{\mathbf{x}^\circ}$ and $S_{\mathbf{y}^\circ}$, expressed in dB, as follows:

- extract the sub-region $\Omega_{\mathbf{x}^\circ}$ as the set of TF points $(m, n) \in \Lambda$ such that $S_{\mathbf{x}^\circ}[m, n] \geq \max(S_{\mathbf{x}^\circ}) - 40\text{dB}$,
- extract $\Omega_{\mathbf{y}^\circ > \mathbf{x}^\circ}$ as the TF points $(m, n) \in \Lambda$ such that $S_{\mathbf{y}^\circ}[m, n] \geq S_{\mathbf{x}^\circ}[m, n]$,
- take the union $\Omega_{\text{raw}} = \Omega_{\mathbf{x}^\circ} \cup \Omega_{\mathbf{y}^\circ > \mathbf{x}^\circ}$;
- smooth Ω_{raw} via standard mathematical morphology techniques [41] by applying a closing followed by an opening with a structuring element of radius 3 and a '+' shape.

⁴<https://gitlab.lis-lab.fr/skmad-suite/tff2020>

⁵<https://freesound.org/>

3) *Performance measures*: to assess numerically the quality of the reconstruction, we use two performance measures. The Source to Distortion Ratio (SDR) [42]

$$\text{SDR}(\mathbf{x}^o, \hat{\mathbf{x}}) = 20 \log_{10} \left(\frac{\|\mathbf{x}^o\|}{\|\hat{\mathbf{x}} - \mathbf{x}^o\|} \right)$$

quantifies the ratio between the energy of the target signal denoted by \mathbf{x}^o and residual error with an estimate $\hat{\mathbf{x}}$. In order to complement this signal-domain evaluation by a TF-domain quality measure [43], [44], we compute the Itakura-Saito (IS) divergence [45] using the discrete Fourier transform \mathcal{F} of the signals, as

$$\text{IS}(\mathbf{x}^o, \hat{\mathbf{x}}) = \sum_f \left(\left| \frac{\mathcal{F}\mathbf{x}^o[f]}{\mathcal{F}\hat{\mathbf{x}}[f]} \right| - \log \left| \frac{\mathcal{F}\mathbf{x}^o[f]}{\mathcal{F}\hat{\mathbf{x}}[f]} \right| - 1 \right)$$

4) *Solvers*: for comparison, we compute 2 baseline solutions and 3 estimates obtained from the proposed approach:

- *Zero-fill*: a baseline method consisting in applying $\mathcal{M}_{\bar{\Omega}}$ to the observed signal \mathbf{z} , i.e., in reconstructing the signal after filling the masked regions by zeros. This is equivalent to take $\lambda_p = \lambda = 1$ in the analytical solution.
- *Interp*: a baseline method consisting in applying a linear interpolation along the frequency axis of the magnitude of observation TF matrix, and by drawing the related phase uniformly at random, as proposed in some industrial software [40].
- *TFF-1*: Time-Frequency Fading method considering the mask as one region (Algo. 1).
- *TFF-P*: Time-Frequency Fading method considering the mask as P sub-regions (Algo. 2).
- *TFF-O*: Time-Frequency Fading method considering the mask as P sub-regions and where λ is chosen so as to maximize the SDR in an oracle way, in order to have an upper bound on the SDR performance.

The parameters used of Rand-EVD are $\epsilon_{\text{EVD}} < 10^{-3}$ and $p_{\text{EVD}} = 0.9999$ and a tolerance $\epsilon = 10^{-5}$ is set for finding the subregions in Algorithm 3.

B. Results

An example of reconstruction is given in Figure 4. One can observe that the spectrograms obtained by the proposed methods are visually very similar to that of the original target signal, except for the largest sub-region on the left. On the contrary, the spectrograms obtained by the baseline methods exhibit large differences.

Detailed results are given in Table I for the SDR and in Table II for the IS divergence. Boldface is used to emphasize the best values among TFF-1, TFF-P, Zero-fill and Interp. In all cases, TFF-1 and TFF-P clearly give SDR values higher than that of baselines Zero-fill, and Interp by several dBs. TFF-1 and TFF-P SDR values are close to the oracle upper bound given by TFF-O, by about 1dB and there is no clear difference between TFF-1 and TFF-P.

In terms of IS divergence, values are much more heterogeneous. TFF-1 and TFF-P still outperform the baseline methods in general. However, the Zero-fill strategy is the best one when the perturbation is impulsive (*clicks* and

finger snaps). In those cases, the perturbation affects a wide frequency range: the Interp strategy is particularly unadapted due to the interpolation along the frequency axis and Zero-fill is giving the lowest (but yet very high) IS values.

Tables I and II display results for a unique realization of the random projections. Running the algorithms several times (100 times in our experiment, for the car and bird signals) gives the same result on average, with a negligible standard deviation (at most $1e-4$ for IS, $1e-11$ for the SDR). This could be expected since the range and eigen-elements provided by the randomized procedure [38] are controlled by theoretical guarantees.

TABLE I
SDR IN DB FOR EACH METHOD AND EACH CONFIGURATION.

Methods		Zero-fill	Interp	TFF-O	TFF-1	TFF-2
beeps	car	20.49	20.17	24.58	23.27	23.26
	train	20.54	20.12	24.51	23.69	23.12
	plane	26.20	25.95	28.23	27.73	27.70
bird	car	12.32	12.08	18.32	17.19	15.30
	train	22.25	21.31	26.02	25.27	23.11
	plane	25.12	24.97	27.40	26.01	25.66
chirps	car	13.64	13.64	21.74	20.03	19.90
	train	13.82	13.58	21.96	18.16	18.28
	plane	18.33	17.90	24.83	22.05	22.32
clicks	car	9.57	9.34	11.82	10.54	11.01
	train	9.21	8.25	11.63	10.49	10.67
	plane	11.45	8.67	13.94	12.69	12.72
finger_snaps	car	8.50	8.29	13.36	11.66	11.48
	train	9.22	8.47	14.64	12.48	12.32
	plane	11.54	10.42	16.41	14.66	14.41
modulations	car	12.93	12.63	17.78	17.40	17.13
	train	17.14	16.89	20.59	19.34	19.56
	plane	21.72	21.53	23.49	22.40	22.86

Figure 5 gives some insights in the case of the bird+car signal and considering only one whole region Ω . Colored points represent the performance obtained by TFF-1, TFF-P, TFF-O and Zero-fill ($\lambda = 1$). The additional points named TFF-E are obtained by extracting the true energy in the subregions from the original target signal, and by using this oracle value as target energy in the estimation of λ . Both curves show similar trends; however, the SDR curve is smooth while the IS curve is more irregular (which originates from our choice of the raw periodogram for spectral estimation; other options such as Welch periodogram yield smoother versions but require additional parameter tuning which we preferred to avoid).

One can see that the λ value estimated by TFF-1 is very close to the TFF-E value, which supports that the proposed energy estimation method is able to accurately estimate the target energy. Both TFF-1 and TFF-E points are close to the optimal TFF-O, in terms of λ value, of high SDR and of low IS divergence, while the Zero-fill points are far away from the optimal.

Due to the processing and the Gabor multipliers diagonalization in high dimension ($L = 16384$ here), the computational

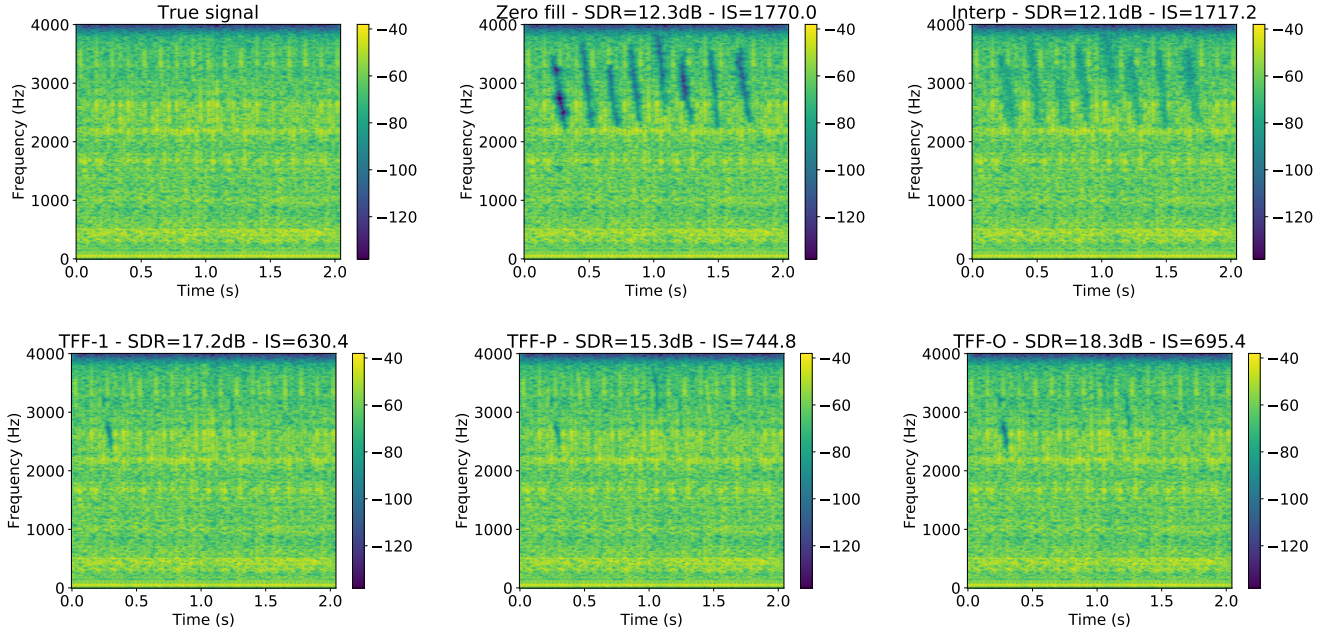
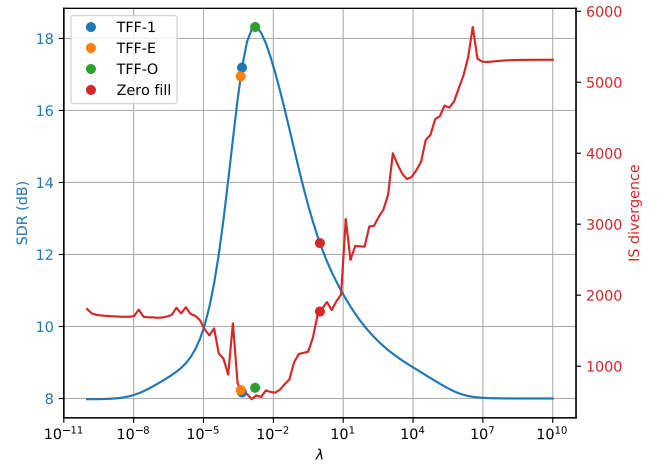


Fig. 4. Bird+car signal: spectrograms of the target signal and of the reconstruction by each method.

TABLE II
IS DIVERGENCES FOR EACH METHOD AND EACH CONFIGURATION.

Methods Signals		zero-fill	Interp	TFF-O	TFF-1	TFF-P
beeps	car	139.4	130.9	63.0	64.9	63.9
	train	270.4	214.1	255.4	138.1	138.1
	plane	452.7	235.4	142.2	151.3	168.3
bird	car	1770.0	1717.2	695.4	630.4	744.8
	train	1870.5	3348.3	1538.3	1943.6	2413.3
	plane	1993.6	1724.1	1383.0	1428.7	1439.5
chirps	car	880.0	868.4	260.3	537.7	613.5
	train	1205.8	1135.2	411.0	1506.8	615.6
	plane	989.1	982.4	450.7	1178.6	736.1
clicks	car	1968.1	2877.0	1685.8	2465.4	2152.7
	train	2392.1	6798.5	2111.5	5092.6	4911.6
	plane	2273.2	8551.7	2507.5	4262.5	4417.5
finger_snaps	car	2961.6	3286.4	1342.5	1902.8	1769.9
	train	3105.5	7668.1	2810.0	3976.7	4263.3
	plane	2629.0	6483.2	2396.7	3665.1	4228.4
modulations	car	832.1	877.4	235.3	282.8	208.2
	train	917.6	1017.7	282.2	349.6	412.7
	plane	1017.8	803.9	580.8	353.5	410.3

complexity of the proposed methods is a key issue that has motivated several modeling and algorithmic contributions in this paper. In [10], many of those contributions were not present: processing a sound of length 8192 was lasting about 300 seconds, and processing longer sounds was both untractable in time and memory. Using TFF-P, processing a twice-longer sound with similar contents takes a bit less than 200 seconds. Running times for TFF-1 and TFF-P are represented in Figure 6, where results are given both for the Hann and Gaussian windows. It shows TFF-P is significantly

Fig. 5. Variation of the performance (SDR and IS) as a function of hyperparameter λ for the bird+car mix, considering a whole region Ω .

faster than TFF-1, while their estimation performance are similar. Indeed, dividing the mask into P sub-regions results in replacing the diagonalization of a large-rank operator with P diagonalizations of lower-rank operators, with significant computational savings. One may also observe that the running times are very different from one case to the other: it is highly dependent on the numerical rank of the Gabor multipliers, which is related to the size of the region Ω .

V. CONCLUSION

In this paper, we have addressed the problem of single- and multiple-region time-frequency fading, formulated as a variational optimization problem. Almost no assumptions are made about the signals to be processed: the only requirement

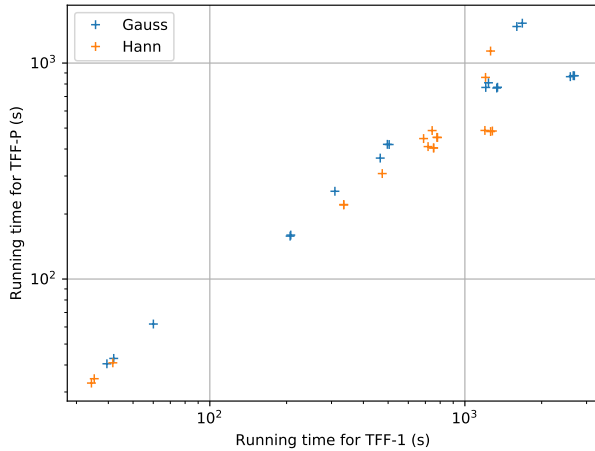


Fig. 6. Computation times (log scale) of TFF-1 versus TFF-P, with one point per mixture, for the Hann and Gaussian windows.

is that the perturbation to be filtered out is well localized in known regions of the time-frequency plane.

We have proposed efficient algorithms based on the diagonalization of Gabor multipliers possibly coupled with fast randomized eigenvalue decomposition. The proposed method is based upon an optimization in the signal domain, avoiding any time-frequency consistency issue (arising when working in the time-frequency domain as mentioned in the introduction). They are easily usable in practice, requiring very few parameters that can be set to some default values. Good reconstruction quality is obtained in experiments on real sounds.

The novel time-frequency fading problem formulation and algorithmic framework may be extended in several directions.

First, the method itself may be studied, improved or extended further. For instance, signals with a duration larger than one or two seconds may be processed easily by segmenting them, processing each segment, and then reconstructing the full-length signal by overlap-add type techniques. From a computational point of view, given the excellent accuracy of the randomized range estimation and EVD [38], we anticipate that substantial savings in terms of computation time can be achieved by lowering the required accuracy, without penalizing the quality of results. Still from a computational point of view, it would be interesting to manage to process each sub-region by considering only its neighborhood, since coefficients that are far away in the time-frequency plane may have an insignificant influence on the sub-region.

Second, our approach is a generic time-frequency filtering technique that may inspire more task-specific variants. It would require to develop new models and optimization schemes that are adapted to the task setting so that the resulting new approaches can be compared to competing methods. An interesting perspective is to integrate *a priori* information about the unknown signals in the optimization problem, e.g. some prior on their spectral shape or on the variations of the coefficients along the time and frequency axes. This can lead to more specific use cases in applications like denoising, source

separation, time-frequency segmentation or inpainting, as well as to developing methods to automatically estimate the mask location. A natural option would be to modify the penalization term, with the risk of losing the spectral formulation developed here that allows rank reduction and further simplifications.

ACKNOWLEDGMENTS

This study originates from a collaboration with ANSYS, which develops sound analysis and design software [40]. We wish to thank ANSYS and more specially S. Molla and H. Lachambre for scientific exchanges and for providing test data and software for comparison purpose. We also thank Emily Bourne for her kind advices to improve the language usage.

REFERENCES

- [1] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *Trans. Sig. Proc.*, vol. 52, no. 7, pp. 1830–1847, July 2004.
- [2] D. Wang, "Time-frequency masking for speech separation and its potential for hearing aid design," *Trends in amplification*, vol. 12, no. 4, pp. 332–353, Dec. 2008.
- [3] P. Lander and E. J. Berbari, "Enhanced ensemble averaging using the time-frequency plane," in *IEEE-SP Proc. Int. Symp. on Time-Freq. and Time-Scale Analysis*, 1994, pp. 241–243.
- [4] F. Hlawatsch, G. Matz, H. Kirchauer, and W. Kozek, "Time-frequency formulation, design, and implementation of time-varying optimal filters for signal estimation," *IEEE Trans. Signal Process.*, vol. 48, no. 5, pp. 1417–1432, 2000.
- [5] L. Stankovic, "On the time-frequency analysis based filtering," *Ann. Télécommun.*, vol. 55, no. 5, pp. 216–225, May 2000.
- [6] B. Boashash, "Time-frequency synthesis and filtering," in *Time-Frequency Signal Analysis and Processing*, pp. 637–691. Elsevier, 2016.
- [7] J. Le Roux and E. Vincent, "Consistent Wiener filtering for audio source separation," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 217–220, Mar. 2013.
- [8] D. Gabor, "Theory of communication. part 1: The analysis of information," *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering*, vol. 93, no. 26, pp. 429–441, Nov. 1946.
- [9] M. R. Portnoff, "Time-frequency representation of digital signals and systems based on short-time Fourier analysis," *IEEE Trans. Acous., Speech Signal Process.*, vol. 28, no. 1, pp. 55–69, 1980.
- [10] A. M. Krémé, V. Emiya, C. Chaux, and B. Torrèsani, "Filtering out time-frequency areas using Gabor multipliers," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, 2020, pp. 5790–5794.
- [11] K. Gröchening, *Foundations of Time-Frequency Analysis*, Birkhäuser, Boston (MA), 2011.
- [12] R. Carmona, W. L. Hwang, and B. Torrèsani, *Practical Time-Frequency Analysis*, vol. 9 of *Wavelet Analysis and Its Applications*, Elsevier, 1998.
- [13] B. Boashash, *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, Academic Press, dec 2015.
- [14] P. Flandrin, *Explorations in Time-Frequency Analysis*, Cambridge University Press, 2018.
- [15] W.C. Knight, R.G. Pridham, and S.M. Kay, "Digital signal processing for sonar," *Proceedings of the IEEE*, vol. 69, no. 11, pp. 1451–1506, 1981.
- [16] P. Balazs, *Regular and Irregular Gabor Multipliers with Application to Psychoacoustic Masking*, Ph.D. thesis, Austrian Academy of Sciences, 2005.
- [17] O. Christensen, *An Introduction to Frames and Bases*, Birkhäuser Basel, 2003.
- [18] S. Qian and D. Chen, "Discrete Gabor transform," *IEEE Transactions on Signal Processing*, vol. 41, no. 7, pp. 2429–2438, jul 1993.
- [19] H. G. Feichtinger and K. Nowak, "A first survey of Gabor multipliers," in *Advances in Gabor Analysis*, H. G. Feichtinger and T. Strohmer, Eds., pp. 99–128. Birkhäuser, 2002.
- [20] W. Kozek and G. Pfander, "Identification of operators with bandlimited symbols," *SIAM J. Math. Anal.*, vol. 37, pp. 867–888, 08 2006.

- [21] H. Bölcskei, “Orthogonal frequency division multiplexing based on offset QAM,” in *Advances in Gabor Analysis*, pp. 321–352, 2003.
- [22] A. Olivero, B. Torrèsani, and R. Kronland-Martinet, “A class of algorithms for time-frequency multiplier estimation,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 8, pp. 1550–1559, 2013.
- [23] H. G. Feichtinger, M. Hampejs, and G. Kracher, “Approximation of matrices by Gabor multipliers,” *IEEE Signal Process. Lett.*, vol. 11, no. 11, pp. 883–886, 2004.
- [24] J. J. Benedetto and G. E. Pfander, “Frame expansions for Gabor multipliers,” *Appl. Comput. Harmon. Analysis*, vol. 20, no. 1, pp. 26–40, 2006, Computational Harmonic Analysis - Part 2.
- [25] M. Dörfler and B. Torrèsani, “On the time-frequency representation of operators and generalized Gabor multiplier approximations,” *J. Fourier Anal. Appl.*, vol. 16, pp. 261–293, 2010.
- [26] K. Schnass and H. G. Feichtinger, “Gabor multipliers, a self-contained survey,” Tech. Rep., Institut für Mathematik der Universität Wien, 2004.
- [27] J. Ramanathan and P. Topiwala, “Time-frequency localization via the Weyl correspondence,” *SIAM J. Math. Anal.*, vol. 24, no. 5, pp. 1378–1393, 1993.
- [28] L. D. Abreu, K. Gröchenig, and J. L. Romero, “On accumulated spectrograms,” in *Trans. Amer. Math. Soc.*, 2016, vol. 368, pp. 3629–3649.
- [29] P. Søndergaard, B. Torrèsani, and P. Balazs, “The linear time frequency analysis toolbox,” *International Journal of Wavelets and Multiresolution Information Processing*, vol. 10, no. 4, pp. 1250032-1–1250032-27, 2012.
- [30] Z. Průša, P. L. Søndergaard, N. Holighaus, Ch. Wiesmeyer, and P. Balazs, “The Large Time-Frequency Analysis Toolbox 2.0,” in *Sound, Music, and Motion*, LNCS, pp. 419–442. Springer International Publishing, 2014.
- [31] D. Arrivault and F. Jalliet, “LTFATPY, a partial python port of the Large Time-Frequency Analysis Toolbox (LTFAT),” available at <https://pypi.org/project/ltfatpy/>, 2018, LabEx Archimede, LIS, I2M, Aix-Marseille Université.
- [32] G. F. Margrave, P. C. Gibson, J. P. Grossman, D. C. Henley, V. Iliescu, and M. P. Lamoureux, “The Gabor transform, pseudodifferential operators, and seismic deconvolution,” *Integrated Comput. Aided Eng.*, vol. 12, pp. 43–55, 2005.
- [33] D. Wang and G. J. Brown, *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*, Wiley-IEEE Press, 2006.
- [34] D. Williamson and D. Wang, “Time-frequency masking in the complex domain for speech dereverberation and denoising,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. PP, pp. 1–1, 04 2017.
- [35] D. T. Stoeva and P. Balazs, “A survey on the unconditional convergence and the invertibility of multipliers with implementation,” in *Sampling – Theory and Applications (A Centennial Celebration of Claude Shannon)*, M. Robinson S. Casey, K. Okoudjou and B. Sadler, Eds. Springer, 2020, to appear, preprint arXiv:1803.00415.
- [36] W. E. Arnoldi, “The principle of minimized iterations in the solution of the matrix eigenvalue problem,” *Quarterly of Applied Mathematics*, vol. 9, no. 1, pp. 17–29, apr 1951.
- [37] C. Lanczos, “An iteration method for the solution of the eigenvalue problem of linear differential and integral operators,” *Journal of Research of the National Bureau of Standards*, vol. 45, no. 4, pp. 255, oct 1950.
- [38] N. Halko, P.-G. Martinsson, and J. A. Tropp, “Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions,” *SIAM J. Math. Anal.*, vol. 53, no. 2, pp. 217–288, 2011.
- [39] K. Ito and R. Nakano, “Optimizing support vector regression hyperparameters based on cross-validation,” in *Proceedings of the International Joint Conference on Neural Networks*, 2003, IEEE.
- [40] ANSYS VRXPERIENCE, “Sound website,” <https://www.ansys.com/fr-fr/products/systems/ansys-vrxperience/sound>.
- [41] R. Van den Boomgard and R. van Balen, “Methods for fast morphological image transforms using bitmapped images,” *Comput. Vision Graph. Image Process.*, vol. 54, no. 3, pp. 252–254, May 1992.
- [42] E. Vincent, R. Gribonval, and C. Févotte, “Performance measurement in blind audio source separation,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, July 2006.
- [43] A. Prodeus and I. Kotvitskyi, “On reliability of log-spectral distortion measure in speech quality estimation,” in *2017 IEEE 4th International Conference Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD)*, oct 2017, IEEE.
- [44] A. De, “Distortion measures for speech coding,” *IETE Technical Review*, vol. 15, no. 4, pp. 251–258, jul 1998.
- [45] F. Itakura and S. Saito, “Analysis synthesis telephony based upon the maximum likelihood method,” in *Proc. 6th of the International Congress on Acoustics*, Los Alamitos, CA, 1968, p. C17–C20.

APPENDIX

A. PROOF OF THE ERROR BOUND (15) We now prove the following theorem.

Theorem. Let $\mathbf{g} \in \mathbb{R}^L$ and Λ generate a Parseval Gabor frame in \mathbb{C}^L . Let $\Omega \subset \Lambda$ be a subset of the time-frequency domain, written as a disjoint union $\Omega = \bigcup_{p=1}^P \Omega_p$. Denote by \mathfrak{M}_{Ω_p} the Gabor multiplier associated with the indicator function of Ω_p , and set

$$\Delta_p = \sum_{q \neq p} \|\mathfrak{M}_{\Omega_p} \mathfrak{M}_{\Omega_q}\|.$$

Let $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_P)$ be a vector of real numbers such that $\lambda_p \in]0, 1[$ for all $p = 1, \dots, P$, and define $\tilde{\lambda}_p = \min_{q \neq p} \lambda_q$. Assume further that

$$\Delta_p < \frac{\lambda_p \tilde{\lambda}_p}{(1 - \lambda_p)(1 - \tilde{\lambda}_p)}$$

Then for all $\mathbf{z} \in \mathbb{C}^L$, if $\mathbf{x}_{\boldsymbol{\lambda}}$ and $\mathbf{x}_{\tilde{\boldsymbol{\lambda}}}^L$ are respectively the solution (8) and the approximate solution (14) of the optimization problem (4), the following bound holds true:

$$\|\mathbf{x}_{\boldsymbol{\lambda}} - \mathbf{x}_{\tilde{\boldsymbol{\lambda}}}^L\|_2 \leq \sum_{p=1}^P \frac{(1 - \tilde{\lambda}_p) \Delta_p}{\lambda_p \tilde{\lambda}_p - (1 - \lambda_p)(1 - \tilde{\lambda}_p) \Delta_p} \|\mathbf{z}\|_2,$$

For the sake of simplicity, we set

$$A_p = (1 - \lambda_p) \mathfrak{M}_{\Omega_p}, \quad \mu_p = \lambda_p / (1 - \lambda_p).$$

The assumption on λ_p then ensures $\|A_p\| < 1$ by Property II.1-(iii). With these notations, we wish to approximate

$$\mathbf{z} - \left(\mathbf{I} - \sum_{q=1}^P A_q \right)^{-1} \sum_{p=1}^P \mu_p A_p \mathbf{z} \approx \mathbf{z} - \sum_{p=1}^P \mu_p (\mathbf{I} - A_p)^{-1} A_p \mathbf{z},$$

in other words

$$\sum_{p=1}^P \mu_p \left[\left(\mathbf{I} - \sum_{q=1}^P A_q \right)^{-1} A_p - (\mathbf{I} - A_p)^{-1} A_p \right] \mathbf{z} \approx \mathbf{0}.$$

Defining $B_p = \sum_{q \neq p} A_q$, we then set

$$\begin{aligned} R_p &= \left(\mathbf{I} - \sum_{q=1}^P A_q \right)^{-1} A_p - (\mathbf{I} - A_p)^{-1} A_p \\ &= \left[\left(\mathbf{I} - (\mathbf{I} - A_p)^{-1} B_p \right)^{-1} - \mathbf{I} \right] (\mathbf{I} - A_p)^{-1} A_p \\ &= \left(\mathbf{I} - (\mathbf{I} - A_p)^{-1} B_p \right)^{-1} (\mathbf{I} - A_p)^{-1} B_p A_p (\mathbf{I} - A_p)^{-1} \end{aligned}$$

since A_p commutes with $(\mathbf{I} - A_p)^{-1}$ by construction. We show below that $\|A_p\| < 1$, $\|B_p\| < 1$, and that under suitable assumptions (specified below as well), $\|(\mathbf{I} - A_p)^{-1} B_p\| < 1$. Then we have

$$\|R_p\| \leq \frac{1}{(1 - \|A_p\|)^2} \frac{\|A_p B_p\|}{1 - \|(\mathbf{I} - A_p)^{-1} B_p\|}.$$

Then notice that

$$\|(\mathbf{I} - A_p)^{-1} B_p\| = \|B_p + (\mathbf{I} - A_p)^{-1} A_p B_p\| \leq \|B_p\| + \frac{\|A_p B_p\|}{1 - \|A_p\|}$$

so that

$$1 - \|(\mathbf{I} - A_p)^{-1} B_p\| \geq \frac{(1 - \|A_p\|)(1 - \|B_p\|) - \|A_p B_p\|}{1 - \|A_p\|}.$$

Coming back to initial notations, we have $\|A_p\| \leq 1 - \lambda_p$ so that $1 - \|A_p\| \geq \lambda_p > 0$. Also, since the sub-regions Ω_q are pairwise disjoint

$$\|B_p\| = \left\| \sum_{q \neq p} A_q \right\| = \left\| \sum_{q \neq p} (1 - \lambda_q) \mathfrak{M}_{\Omega_q} \right\| \leq 1 - \min_{q \neq p} \lambda_q,$$

therefore

$$1 - \|B_p\| \geq \tilde{\lambda}_p, \quad \text{where} \quad \tilde{\lambda}_p = \min_{q \neq p} \lambda_q > 0.$$

Besides,

$$\|A_p B_p\| = (1 - \lambda_p) \left\| \sum_{q \neq p} (1 - \lambda_q) \mathfrak{M}_p \mathfrak{M}_q \right\| \leq (1 - \lambda_p)(1 - \tilde{\lambda}_p) \Delta_p,$$

where we have set $\Delta_p = \sum_{q \neq p} \|\mathfrak{M}_{\Omega_p} \mathfrak{M}_{\Omega_q}\|$. Plugging this expression into the lower bound of $1 - \|(\mathbf{I} - A_p)^{-1} B_p\|$ yields

$$1 - \|(\mathbf{I} - A_p)^{-1} B_p\| \geq \frac{\lambda_p \tilde{\lambda}_p - (1 - \lambda_p)(1 - \tilde{\lambda}_p) \Delta_p}{1 - \|A_p\|}.$$

From now on we assume that

$$\Delta_p < \frac{\lambda_p \tilde{\lambda}_p}{(1 - \lambda_p)(1 - \tilde{\lambda}_p)}$$

so that the lower bound is strictly positive. Putting things together we then obtain

$$\|R_p\| \leq \frac{1 - \lambda_p}{\lambda_p} \frac{(1 - \tilde{\lambda}_p) \Delta_p}{\lambda_p \tilde{\lambda}_p - (1 - \lambda_p)(1 - \tilde{\lambda}_p) \Delta_p}.$$

The bound (15) then follows from

$$\|\mathbf{x}_\lambda - \mathbf{x}_\lambda^L\|_2 \leq \sum_p \frac{\lambda_p}{1 - \lambda_p} \|R_p\| \|\mathbf{z}\|_2.$$

B. PROOF OF GRADIENT COMPUTATION (5)

By definition of the Frobenius norm $\|\mathbf{F}\|^2 = \text{Tr}(\mathbf{F}^H \mathbf{F})$, and using the properties of the trace, we have for any $\Omega \subset \Lambda$, $\|\mathcal{V}\mathbf{x}\|_\Omega^2 = \text{Tr}(\mathbf{x}^H \mathcal{V}^H 1_\Omega \mathcal{V} \mathbf{x}) = \text{Tr}(\mathbf{x}^H \mathfrak{M}_\Omega \mathbf{x}) = \text{Tr}(\mathbf{x}^T \mathfrak{M}_\Omega \mathbf{x})$, since \mathbf{x} is a real vector. Notice in addition that since Ω is symmetric with respect to the zero frequency, \mathfrak{M}_Ω is a real matrix, which is therefore symmetric (since \mathfrak{M}_Ω is self-adjoint).

The objective function may be written as

$$\begin{aligned} f_\lambda(\mathbf{x}) &= \|\mathcal{V}\mathbf{x} - \mathcal{V}\mathbf{z}\|_\Omega^2 + \sum_{p=1}^P \lambda_p \|\mathcal{V}\mathbf{x}\|_{\Omega_p}^2 \\ &= \text{Tr}((\mathbf{x} - \mathbf{z})^T \mathfrak{M}_\Omega (\mathbf{x} - \mathbf{z})) + \sum_{p=1}^P \lambda_p \text{Tr}(\mathbf{x}^T \mathfrak{M}_{\Omega_p} \mathbf{x}). \end{aligned}$$

Using the derivation formula $\nabla_{\mathbf{x}} \text{Tr}(\mathbf{x}^H A \mathbf{x}) = (A + A^T) \mathbf{x}$ yields the result, since \mathfrak{M}_Ω is symmetric.



Marina Krémé obtained a Master's degree in Mathematics and Applications in Probability and Statistics from the University of Abidjan, Côte d'Ivoire, 2015. She obtained a second Master's degree in Mathematics and Computer Science, Statistics, Signals and Health at Aix-Marseille University, France, 2017. She is currently finishing her PhD thesis at the Institut de Mathématiques de Marseille (I2M, UMR-CNRS 7373) and the Laboratoire d'Informatique et des Systèmes (LIS, UMR-CNRS 7020). Her research interests focus on audio signal processing, convex optimization and Machine learning.



Valentin Emiya graduated from Telecom Bretagne, France (2003) and received the M.Sc. degree in Acoustics, Signal Processing and Computer Science Applied to Music (ATIAM) at Ircam, France (2004). He received a PhD in Signal Processing at Telecom ParisTech, France (2008). He was then a post-doctoral researcher with the METISS group at Inria, Rennes, France. Since 2011 as an assistant professor in computer science at Aix-Marseille Univ., France, he has been a member of machine learning team QARMA at Laboratoire d'Informatique et des Systèmes.

His research interests focus on machine learning and signal processing.



Caroline Chaux (S'05-M'07-SM'13) received the engineering degree in telecommunications from the Institut des Sciences de l'Ingénieur de Toulon et du Var (ISITV), France, and the DEA degree in Signal and Digital Communications from the Université de Nice Sophia-Antipolis, France in 2003. In 2006, she then received the PhD degree in signal and image processing from University Paris-Est (Laboratoire d'Informatique Gaspard Monge UMR-CNRS 8049), France, 2006. In 2006-07, she was post-doctoral fellow with the ARIANA research group (INRIA Sophia-Antipolis Méditerranée) before being appointed the same year by CNRS as research scientist in the Laboratoire d'Informatique (UMR-CNRS 8049) of the University Paris-Est. In 2012, she moved to the Institut de Mathématiques de Marseille (UMR-CNRS 7373) of Aix-Marseille Université.



Bruno Torrèsani was born in Marseille, France, in 1961. He received the Ph.D. degree in theoretical physics from Université de Provence, France, in 1986, and the Habilitation degree from Université de la Méditerranée, France, in 1993. From 1989 to 1997, he was a research fellow at the French Centre National de la Recherche Scientifique (CNRS). He currently holds a Professor position at the Mathematics Department and Institut de Mathématiques de Marseille, Aix-Marseille University, Marseille, France. His current research interests are mainly in

the domains of mathematical signal processing, with emphasis on harmonic analysis and probability based approaches, and applications in audio signal processing and brain imaging and computational biology. His other domains of expertise are mathematical physics, and electromagnetic scattering theory. He has authored 3 books, and more than 70 articles in international journals and 80 in international conference proceedings. B. Torrèsani is currently an Associate Editor of Applied and Computational Harmonic Analysis, the International Journal of Wavelets and Multiresolution Information Processing, and the Journal of Signal, Image and Video Processing.

Time-frequency fading algorithms based on Gabor multipliers

Part II: Supplementary material

A. Marina Krémé, Valentin Emiya, Caroline Chaux, *Senior member, IEEE*, and Bruno Torr  sani

Summary. This note provides additional information and results for the paper *Time-frequency fading algorithms based on Gabor multipliers* [1]. The paper is mainly about time-frequency filtering, more precisely time-frequency filter design. The main contribution is a new method for canceling or attenuating signal "components" that are essentially localized in pre-specified regions in the time-frequency plane. This method departs from time-frequency masking and is formulated as a variational approach that focuses on time-frequency localization properties. Although the problem possesses a closed-form solution, the latter leads to important computational costs for real-life signals, which is addressed via dimension reduction through random projections.

COMPARISON WITH CLASSICAL TIME-FREQUENCY FILTERS AND ORACLE APPROACHES

The time-frequency fading (TFF) problem may be described as follows: given a region Ω in the time-frequency plane, construct a linear operator that filters out that region, very much like a classical time-invariant stop-band filter that would filter out a prescribed frequency band. In TFF, the region Ω is fixed. TFF depends on a parameter λ (which may be scalar, or a vector when Ω can be split as a disjoint union of connected sub-regions). [1] also provides a procedure for the corresponding parameter estimation, based upon an energy criterion.

In [1] the method is evaluated on a simulated problem where a wideband signal is perturbed by an additive *perturbation signal* with known time-frequency localization. Results obtained using two variants of TFF are compared with baseline methods: binary masking (called zero-fill) and amplitude interpolation within Ω followed by phase assignment. They are also compared with an oracle version, which addresses the tuning of the parameter λ . Since the determination of the region Ω is not part of the algorithm, it is not addressed in [1].

It is also interesting to compare results with more traditional time-frequency filtering methods, such as Wiener-type filtering, or oracle approaches such as described in [2], even though they do not address the same problem. There, the reference and perturbation signals are supposed to be known (or at least their spectrograms), and an optimal filter can then be searched for. Since only redundant time-frequency transforms are considered, the Wiener filter is not tractable for real-life signals, and is replaced with heuristics called pseudo-Wiener filters.

Starting from the observation model considered in [1]

$$\mathbf{z} = \mathbf{x}^o + \mathbf{y}^o, \quad (1)$$

we considered three versions of the pseudo Wiener filter, which may be formulated as three different Gabor multipliers $\mathcal{V}^{-1}\mathbf{m}\mathcal{V}$ with different masks

$$\text{v1: } \mathbf{m} = \frac{|\mathcal{V}\mathbf{x}^o|^2}{|\mathcal{V}\mathbf{z}|^2}; \quad \text{v2: } \mathbf{m} = \frac{|\mathcal{V}\mathbf{x}^o|^2}{|\mathcal{V}\mathbf{x}^o|^2 + |\mathcal{V}\mathbf{y}^o|^2}; \quad \text{v3: } \mathbf{m} = \Re\left(\frac{\mathcal{V}\mathbf{x}^o}{\mathcal{V}\mathbf{z}}\right). \quad (2)$$

Corresponding evaluation results for the Signal to Distortion Ratio (SDR) and for the Itakura-Saito (IS) divergence are given in Table I .

This Table also reports results obtained using then Oracle approaches derived in [2], where an optimal mask (real or binary) is obtained from the knowledge of the reference and perturbation signal (we adapted the function `bss_nearopt_monomask` from the corresponding toolbox¹ to use time-frequency tools used throughout our simulations).

¹http://bass-db.gforge.inria.fr/bss_oracle/

TABLE I
SDR AND IS DIVERGENCES FOR 3 VARIANTS OF PSEUDO-WIENER FILTERING AND EACH MIXTURE, AND ORACLE (REAL AND BINARY MASKS).

		SDR					IS				
		Wiener v1	Wiener v2	Wiener v3	Oracle-R	Oracle-B	Wiener v1	Wiener v2	Wiener v3	Oracle-R	Oracle-B
Signals	Methods										
beeps	car	21.73	23.36	27.57	24.57	23.21	97.40	70.24	34.67	62.10	84.85
	train	24.08	24.12	27.57	25.15	23.95	138.90	128.17	64.60	116.56	145.24
	plane	28.30	28.21	32.24	29.50	27.93	120.65	153.00	56.29	113.55	230.01
bird	car	16.98	18.04	20.72	18.63	18.11	684.46	538.52	440.02	445.26	642.76
	train	25.78	26.50	29.97	27.74	26.09	1287.28	1220.07	669.11	952.86	1279.37
	plane	27.65	27.97	31.70	29.27	27.72	1152.24	1149.71	560.97	982.45	1211.67
chirps	car	17.27	18.53	21.54	19.27	18.55	507.13	483.38	227.51	526.44	426.87
	train	17.50	19.36	22.20	20.02	19.37	757.89	608.32	324.08	456.70	567.52
	plane	21.25	22.92	25.91	23.84	22.45	657.97	557.39	267.53	437.68	844.20
clicks	car	16.93	18.23	20.72	18.83	17.98	648.57	616.22	305.18	496.65	566.17
	train	18.57	19.46	22.16	20.13	19.25	1181.36	1005.64	620.90	894.55	1154.28
	plane	17.83	18.84	21.20	19.50	19.20	1155.64	1134.44	585.18	901.68	1145.20
finger_snaps	car	14.80	16.17	18.77	16.66	15.98	1130.57	769.52	434.32	1282.38	750.16
	train	16.59	17.69	20.20	18.30	17.62	1580.42	1166.95	723.17	928.73	1180.99
	plane	18.59	19.65	21.95	20.30	19.61	1565.76	1414.71	694.78	1183.58	1302.65
modulations	car	16.19	17.86	20.44	18.49	17.47	406.79	372.72	152.43	229.70	330.02
	train	19.10	21.00	24.89	21.93	21.06	558.59	493.60	195.05	311.36	545.04
	plane	23.49	24.73	28.15	25.93	24.34	577.41	523.55	209.00	381.55	568.59

As may be expected, the best results are obtained with variant 3 of the pseudo-Wiener filter, which uses more information than others. The real oracle mask outperforms the other two pseudo-Wiener filters, while the oracle binary mask yields weaker performances (which could also be expected).

Even though these results are interesting by themselves (we thank the anonymous reviewer who brought this point to our attention) we believe a direct comparison with results from TFF is not fully relevant, because of the following two points

- Pseudo-Wiener filters and Oracle BSS approaches use information that is not available to TFF, and can therefore be expected to outperform TFF (which is indeed the case).
- However, they are based upon real or binary time-frequency masking, which is a heuristics, while TFF builds the time-frequency filter as the true solution (up to the precision of our approximations, which are controlled) of an optimization problem.

CONSTRUCTION OF MASKS

In the numerical experiments of [1], an ad-hoc procedure is used for constructing the region Ω . The latter uses the signals \mathbf{x}^o and \mathbf{y}^o of the model (1), and operates a series of thresholding and morphological operations. The goal in [1] was to have a simple procedure for mask generation, no particular attention was paid to its optimization, which was not a goal of the work. Several variants have been used in pilot studies, yielding comparable results.

An alternative could have been to use the oracle binary mask proposed in [2], our results show that the latter is quite irregular (many isolated points) and requires significant post-processing (like the raw mask of [1]), at least comparable with post-processing done in our work [1]. Designing specific variational approaches exploiting the characteristics of the chosen Gabor system would be an interesting development.

REFERENCES

- [1] A.-M. Kr  m  , V. Emiya, C. Chaux and B. Torr  sani, Time-frequency fading algorithms based on Gabor multipliers, IEEE Journal of Selected Topics in Signal Processing, 2020.

[2] E. Vincent, R. Gribonval and M.D. Plumbley, Oracle estimators for the benchmarking of source separation algorithms, *Signal Processing* 87(8), p. 1933-1950, 2007