



Filtering out time-frequency areas using Gabor multipliers

Ama Marina Krémé, Valentin Emiya, Caroline Chaux, Bruno Torresani

► To cite this version:

Ama Marina Krémé, Valentin Emiya, Caroline Chaux, Bruno Torresani. Filtering out time-frequency areas using Gabor multipliers. ICASSP: 45th International Conference on Acoustics, Speech, and Signal Processing, May 2020, Barcelone, Spain. hal-02456153

HAL Id: hal-02456153

<https://hal.science/hal-02456153>

Submitted on 27 Jan 2020

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 destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

FILTERING OUT TIME-FREQUENCY AREAS USING GABOR MULTIPLIERS

A. Marina Krémé^{1,2}, Valentin Emiya², Caroline Chaux¹ and Bruno Torrèsani¹

¹ Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France

² Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France

ABSTRACT

We address the problem of filtering out localized time-frequency components in signals. The problem is formulated as a minimization of a suitable quadratic form, that involves a data fidelity term on the short-time Fourier transform outside the support of the undesired component, and an energy penalization term inside the support. The minimization yields a linear system whose solution can be expressed in closed form using Gabor multipliers. We provide an analysis of the solution and its approximation by truncated eigenvalue expansion, and illustrate its performance on synthetic mixtures of audio signals. The proposed approach outperforms approaches that are routinely used in applications.

Index Terms— Time frequency filtering, Gabor transform, Gabor multipliers.

1. INTRODUCTION

Processing signals whose frequency content changes over time is widely addressed using time-frequency (TF) representations [1], mainly the short-time Fourier transform (STFT) and its discrete counterpart the Gabor transform [2] for numerical calculations. Audio source separation is an active research field where such representations are processed [3].

A key challenge is that of filtering out localized TF components. Methods based on ridge extraction, reassignment and synchrosqueezing [4, 5, 6] model and subtract the components. One may also consider the distribution of zeros in the TF plane to locate and remove components [7]. However, those methods are designed to preserve the quality of the residual signal. Another strategy consists in inpainting the spectrogram coefficients related to the removed components, using nonnegative matrix factorization [8, 9] or linear interpolation with random phases [10]. The resulting spectrograms suffer from being inconsistent [11], which often cannot be fixed by phase reconstruction methods [12].

In this paper, we investigate a non-parametric approach to filtering out a localized TF components by reconstructing a signal with STFT coefficients that both fit the original signal outside the TF support of the filtered components and by penalizing the energy inside their TF support. In section 2, we propose a simple and generic formulation of the optimization

problem. The signal itself is directly reconstructed, avoiding any consistency issue. We show in section 3 that the solution is obtained in closed form. It involves the so-called Gabor multipliers [13, 14] and their spectral decomposition, which are computed independently of the hyperparameters so that they can be tuned in a computationally-efficient way. Numerical experiments proposed in section 4 reveal the detailed behavior of the methods, including the TF localization of the Gabor multipliers' eigenvectors, and the quality of reconstructed target and interference sources.

2. PROBLEM STATEMENT

Let \mathcal{T} denote a generic time-frequency transform, mapping any $\mathbf{x} \in \mathbb{C}^L$ to $\mathcal{T}\mathbf{x} \in l^2(\Lambda)$, Λ being the discrete time-frequency domain. Our goal is to filter out signal components located in a given region of the TF plane, in the following setting: we assume that the observed signal \mathbf{x}_0 is of the form $\mathbf{x}_0 = \mathbf{x}_{\text{ref}} + \mathbf{x}_{\text{per}}$, i.e. the sum of a target signal \mathbf{x}_{ref} , and a perturbation \mathbf{x}_{per} . While no assumption is made on the target signal, we assume further that the STFT of \mathbf{x}_{per} is strongly concentrated in a known region $\Omega \subset \Lambda$. We denote by $\bar{\Omega} = \Lambda \setminus \Omega$ the complement region. Estimating \mathbf{x}_{ref} from the observation \mathbf{x}_0 and from the knowledge of support Ω is a specific denoising or source separation problem which we formulate as

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathbb{C}^L}{\operatorname{argmin}} \quad \|\mathcal{T}\mathbf{x} - \mathcal{T}\mathbf{x}_0\|_{\bar{\Omega}}^2 + \lambda \|\mathcal{T}\mathbf{x}\|_{\Omega}^2. \quad (1)$$

where $\|\mathbf{y}\|_{\Omega}^2 := \sum_{k \in \Omega} |\mathbf{y}[k]|^2$ and $\lambda > 0$. The first term of the objective function in (1) is a data fidelity term that enforces the STFT of the estimated signal to fit that of the observation outside Ω . The second term controls its energy in Ω , and the regularization parameter λ controls the trade-off between the two terms.

Here and throughout the paper, we will focus on the Gabor transform although problem (1) can be solved in a more general case. We first introduce some definitions and notations. Let a and b denote the time and frequency sampling period, assumed to be divisors of L , and set $N = L/a$, $M = L/b$. N and M are respectively the numbers of time shifts and frequency shifts. The TF domain is the lattice

$\Lambda = \mathbb{Z}_M \times \mathbb{Z}_N$. For a window function $\mathbf{g} \in \mathbb{R}^L$, the Gabor atom \mathbf{g}_{mn} at TF point $(m, n) \in \Lambda$ is defined by $\mathbf{g}_{mn}[l] = \mathbf{g}(l - na)e^{2i\pi mbl/M}$, $\forall l \in \mathbb{Z}_L$. The Gabor transform $\mathcal{V}_{\mathbf{g}}\mathbf{x} \in \mathbb{C}^{M \times N}$ of \mathbf{x} is defined by

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

Given the adjoint $\mathcal{V}_{\mathbf{g}}^*$ of $\mathcal{V}_{\mathbf{g}}$, also called *synthesis operator*, the *frame operator* $\mathcal{S} = \mathcal{V}_{\mathbf{g}}^* \mathcal{V}_{\mathbf{g}}$ is bounded, self-adjoint and semi-positive definite. If $\mathcal{S} > 0$, \mathcal{S} is invertible, which permits to reconstruct any $\mathbf{x} \in \mathbb{C}^L$ from its Gabor coefficients. Of particular interest are the so-called Parseval frames, for which $\mathcal{S} = I$, i.e., $\mathcal{V}_{\mathbf{g}}^*$ is a left inverse of $\mathcal{V}_{\mathbf{g}}$. For the sake of simplicity, we focus on this case in this paper.

3. ANALYTICAL RESOLUTION

Using Gabor multipliers introduced in section 3.1, the analytical solution of problem (1) is established in section 3.2.

3.1. Gabor multipliers

Gabor multipliers are linear operators on \mathbb{C}^L defined by pointwise multiplication with a TF transfer function called a *Gabor mask*, in the Gabor coefficients domain. Denoting by \mathbf{m} the Gabor mask, we will also denote for simplicity by \mathbf{m} the operator of pointwise multiplication by \mathbf{m} .

Definition 3.1 *The Gabor multiplier associated to (\mathbf{g}, Λ) with mask \mathbf{m} is defined by $\mathfrak{M}_{\mathbf{m}} = \mathcal{V}_{\mathbf{g}}^* \mathbf{m} \mathcal{V}_{\mathbf{g}}$, i.e.,*

$$\mathfrak{M}_{\mathbf{m}}\mathbf{x} = \sum_{m,n} \mathbf{m}[m, n] \langle \mathbf{x}, \mathbf{g}_{m,n} \rangle \mathbf{g}_{m,n}.$$

Gabor multipliers have been studied extensively (see [15, 16] and references therein). We recall below some important properties that will be of interest in the sequel.

Properties 3.1

- (i) *If \mathbf{m} is real-valued then $\mathfrak{M}_{\mathbf{m}}$ is self-adjoint. Then there is 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}, Λ) generates a tight Gabor frame.*
- (iii) *If $\mathbf{m} \in \mathbb{C}^{M \times N}$, then $\mathfrak{M}_{\mathbf{m}}$ defines a bounded operator with operator norm $\|\mathfrak{M}_{\mathbf{m}}\|_{op} \leq C \|\mathbf{m}\|_{\infty}$, where C is a constant. In particular, if \mathbf{g} and Λ generate a Parseval frame, then $\|\mathfrak{M}_{\mathbf{m}}\|_{op} \leq \|\mathbf{m}\|_{\infty}$.*

3.2. Analytical solution

Let $f(\mathbf{x}) = \|\mathcal{V}_{\mathbf{g}}\mathbf{x} - \mathcal{V}_{\mathbf{g}}\mathbf{x}_0\|_{\Omega}^2 + \lambda \|\mathcal{V}_{\mathbf{g}}\mathbf{x}\|_{\Omega}^2$ be the objective function in (1) for $\mathcal{T} = \mathcal{V}_{\mathbf{g}}$. f is a quadratic form and therefore its minimization results in a linear system. For simplicity, we denote by \mathfrak{M}_{Ω} and $\mathfrak{M}_{\bar{\Omega}}$ the multipliers associated with the indicator functions $\mathbb{1}_{\Omega}$ and $\mathbb{1}_{\bar{\Omega}}$ considered as Gabor masks. We obtain

$$\nabla f(\mathbf{x}) = 2(\mathfrak{M}_{\bar{\Omega}} + \lambda \mathfrak{M}_{\Omega})\mathbf{x} - 2\mathfrak{M}_{\bar{\Omega}}\mathbf{x}_0.$$

Since $\mathcal{V}_{\mathbf{g}}^* \mathcal{V}_{\mathbf{g}} = I$, we have $\mathfrak{M}_{\bar{\Omega}} = I - \mathfrak{M}_{\Omega}$ and then $\mathfrak{M}_{\bar{\Omega}} + \lambda \mathfrak{M}_{\Omega} = [I + (\lambda - 1)\mathfrak{M}_{\Omega}]$. According to property 3.1 (iii), $(\mathfrak{M}_{\bar{\Omega}} + \lambda \mathfrak{M}_{\Omega})$ is invertible if $0 < \lambda < 2$. Then:

$$\nabla f(\mathbf{x}^*) = 0 \iff \mathbf{x}_{\lambda}^* = (\mathfrak{M}_{\bar{\Omega}} + \lambda \mathfrak{M}_{\Omega})^{-1} \mathfrak{M}_{\bar{\Omega}} \mathbf{x}_0.$$

As the mask is real valued and using properties 3.1(i) and 3.1(iii), there is an unitary matrix U and a diagonal matrix $D = \text{diag}(\sigma_1, \dots, \sigma_L)$, $\sigma_1 \geq \dots \geq \sigma_L$ such that $\mathfrak{M}_{\Omega} = U D U^{-1}$. Setting $\gamma_l = \frac{\lambda \sigma_l}{1 - (1 - \lambda) \sigma_l}$, we then have:

$$\mathbf{x}_{\lambda}^* = \mathbf{x}_0 - U \text{diag}(\gamma_1, \dots, \gamma_L) U^{-1} \mathbf{x}_0. \quad (2)$$

While the above $L \times L$ matrices can be very large, the Gabor multipliers under consideration act on much lower dimensional subspaces. The eigenvalues σ_l being sorted in decreasing order, sequence $\{\gamma_l, l = 1, \dots, L\}$ is decreasing too, which can be used for truncation purpose. Let be \mathbf{x}_K^* be the solution obtained when only the $K < L$ largest eigenvalues are retained, i.e. $\mathbf{x}_K^* = U_{1:K} \text{diag}(\gamma_{1:K}) U_{1:K}^{-1} \mathbf{x}_0$. The truncation error is controlled by

$$\begin{aligned} \|\mathbf{x}_{\lambda}^* - \mathbf{x}_K^*\| &= \|\text{diag}(0, \dots, 0, \gamma_{K+1}, \dots, \gamma_L) U^{-1} \mathbf{x}_0\|_{\mathbb{C}^L} \\ &\leq \gamma_{K+1} \|U^{-1} \mathbf{x}_0\|_{\mathbb{C}^L} \leq \gamma_{K+1} \|\mathbf{x}_0\| \end{aligned} \quad (3)$$

4. EXPERIMENTS

We illustrate the approach on a synthetic mixture of two real audio signals. We first describe the signals, before analyzing the action of Gabor multipliers and discussing results.

4.1. Experimental setting

The sounds were provided by ANSYS VRXPERIENCE Sound Analysis and Specification¹. The target signal \mathbf{x}_{ref} is a car engine sound and the perturbation signal \mathbf{x}_{per} is a birdsong. Both signals have been sampled at $f_s = 8000$ Hz and are $L = 8192$ samples long. The Gabor transform for each of these signals is calculated with a Hann window of length 128, the time-frequency lattice parameters are set to $a = 32$ and $b = 512$, generating a 256×256 TF matrix. The observation is a linear combinations of these two signals, as shown in Fig. 1. The spectrograms of the engine sound and birdsong are displayed in the first row of Fig. 4.

¹<https://www.ansys.com/fr-fr/products/systems/ansys-vrxperience/sound>

The goal is to filter out the birdsong. To this end, a mask \mathbf{m} was constructed as the indicator function of a region Ω matching the six high frequency components as shown in Fig. 1. The mask is constructed from spectrograms of both sources as:

$$\mathbf{m}_{ij} = \begin{cases} 1 & \text{if } |\mathcal{V}_{\mathbf{g}}\mathbf{x}_{\text{per}}(i, j)| < 2|\mathcal{V}_{\mathbf{g}}\mathbf{x}_{\text{ref}}(i, j)|. \\ 0 & \text{else} \end{cases}$$

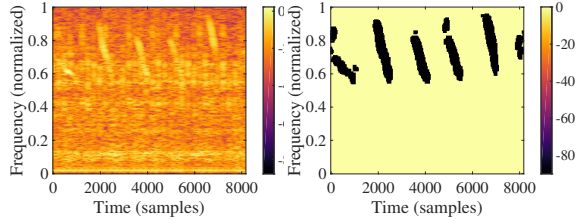


Fig. 1. Sum of an engine's sound and a birdsong: observations (left) and binary mask (right).

The closed-form solution (2) depends on the regularization parameter $0 < \lambda < 2$, which is adjusted using the following strategy. We choose the optimal λ as the value for which the energy of the reconstructed signal $\|\mathcal{V}_{\mathbf{g}}\mathbf{x}_{\lambda}^*\|_{\Omega}^2$ within the region Ω matches the energy $E = \|\mathcal{V}_{\mathbf{g}}\mathbf{x}_0\|_{\Omega'}^2$, in another region Ω' similar to and disjoint from Ω . Ω' is grossly chosen by hand here and this process may be automated if needed. Such tuning of λ is computationally efficient: since the spectral decomposition of the Gabor multiplier does not depend on λ , testing each value of λ mainly requires the truncated matrix-vector multiplications to build \mathbf{x}_K^* and the computation of its Gabor transform.

4.2. Gabor multiplier's spectral decomposition

The Gabor multiplier eigenvalues σ_l and coefficients γ_l are shown in Fig. 2. Only the first 4000, which are above numerical accuracy, are displayed. γ values below 10^{-5} were truncated, leading to a $K = 1696$ -dimensional subspace.

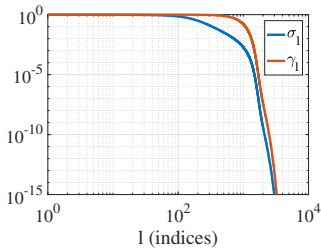


Fig. 2. Gabor multiplier's eigenvalues σ_l and coefficients γ_l .

Interestingly, the spectral decomposition of the Gabor multiplier \mathcal{M}_{Ω} tends to separate contributions from the various connected components. For example, we display in Fig. 3

the spectrograms of the 100th and 1500th eigenvectors, together with the time and spectral representations. The first one appears to be sharply localized in the region of one of the component, while the other one is closer to its boundary. This is in agreement with the usual behavior of Gabor multiplier eigenfunctions. However we find quite interesting the fact that multipliers associated with disconnected regions tend to generate eigenfunctions localized in the connected components.

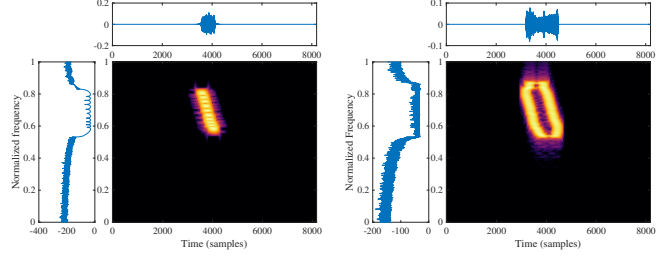


Fig. 3. Waveform, spectrogram and spectrum of the 100th (left) and 1500th (right) eigenvectors of the Gabor multiplier.

4.3. Comparative reconstruction results

We compare here results obtained with the proposed approach, hereafter termed RedEnergy, with results obtained using two approaches commonly used in industrial applications:

- **ZerVal** (zero values): set to zero coefficients within the region Ω before inverting the Gabor transform, i.e. apply the Gabor multiplier \mathcal{M}_{Ω} .
- **RandVal** (random values, as implemented in the commercial software SAS [10]): estimate the Gabor coefficient modulus within Ω by linear interpolation along the frequency axis, then turn to complex coefficients by generating random, uniformly distributed phases, before inverting the Gabor transform.

The spectrograms of the target signal and the perturbation signal reconstructed by RedEnergy, ZerVal and RandVal respectively are presented in Fig. 4. Regarding the engine sound (left column), the spectrogram of the RedEnergy reconstruction (Fig. 4, 2nd row) is visually very close to the original, while significant differences can be seen with the outputs of ZerVal and RandVal (respectively 3d and 4th rows of Fig. 4), where the birdsong is still present. Regarding the birdsong (right column), RedEnergy also outperforms the other two methods. Note that the birdsong reconstructions are of poorer quality because the low frequency part of the chirps was not included in the domain Ω .

Quantitative assessment of reconstruction quality for \mathbf{x}_{ref} was measured using a signal-to-noise ratio (SNR), SNR

values are displayed at the top of spectrograms in Fig. 4. The evolution of SNR as a function of λ is displayed in Fig 5, together with SNR values obtained with ZerVal and RandVal. The plot shows that the SNR value for the optimal λ value is significantly better than those obtained by ZerVal and RandVal, which confirms the visual impression.

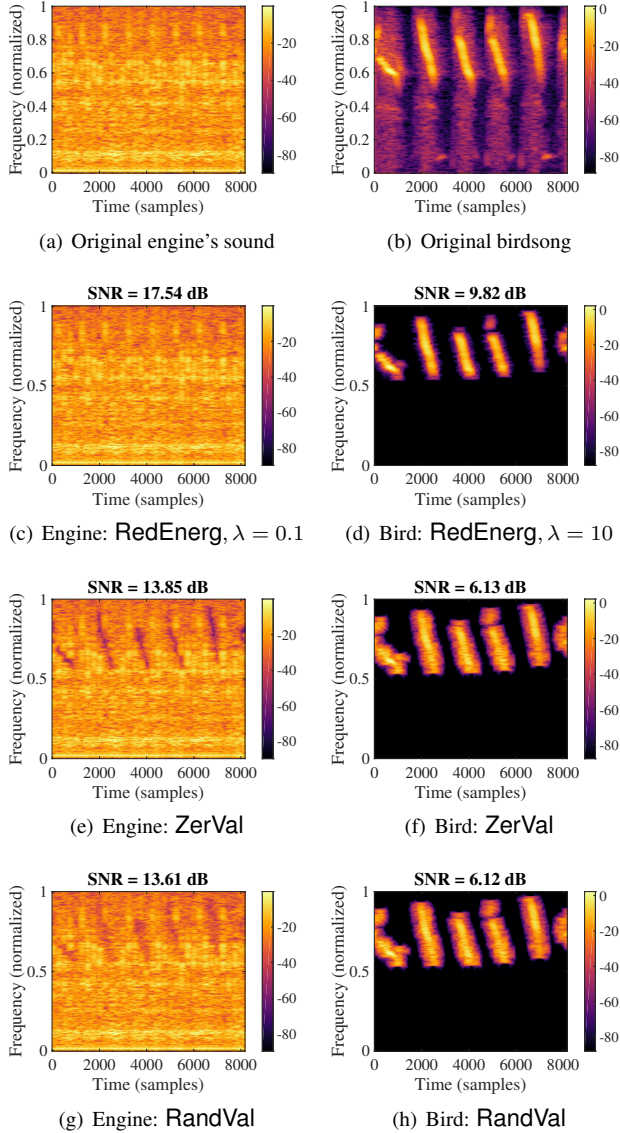


Fig. 4. Spectrogram of the reconstructed signals. Left column: Engine's sound. Right column: Birdsong.

The performance assessment of the proposed method is confirmed when extending the same experiments to other sounds, downloaded from Freesound². Target sounds include a car engine, a train and an airplane while 4 perturbation signals well located in the TF plane are used: some beeps, a finger snap, clicks and another birdsong. For each mixture,

²<https://freesound.org>

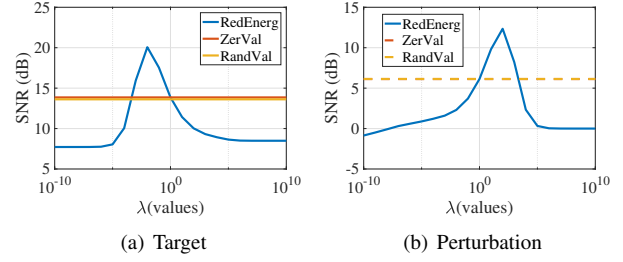


Fig. 5. SNR in decibels (dB) for the three solvers.

the SNRs calculated between engine sound and the signal estimated by the three solvers are summarized in Table. 1 and show the superiority of the proposed method.

From a computational viewpoint, all methods were implemented in MATLAB on a macOS system with 2.3 GHz Intel core i5. The most time-consuming part is the diagonalization of the multiplier (about 306 seconds), which is done only once. The computation of the solution then being very fast: the adjustment of the regularization parameter λ , that involves many evaluations of the solution, took about 114 seconds.

	car	train	aircraft	
RedEnergy	25.05	23.88	26.80	Beeps
ZerVal	24.23	16.52	24.73	
RandVal	24.90	16.15	24.73	
RedEnergy	18.88	22.49	20.72	Finger snap
ZerVal	17.18	20.70	16.59	
RandVal	16.92	16.17	15.93	
RedEnergy	21.67	20.97	17.96	Clicks
ZerVal	17.61	13.98	12.33	
RandVal	10.57	5.60	11.07	
RedEnergy	18.01	21.52	20.04	Birdsong
ZerVal	17.4	21.52	19.71	
RandVal	17.13	21.28	19.56	

Table 1. SNRs for several targets and perturbations.

5. CONCLUSION

We have addressed the problem of estimating a target signal, with no assumption on its contents, when perturbed by an additive signal that is well located in a region Ω of the TF plane. We have proposed an optimization problem in which the energy in Ω is controlled. It admits an analytical solution, which provides the estimated signal directly, without suffering from TF consistency issues. The underlying Gabor multipliers eigenvectors show interesting localization properties and the proposed method outperforms some industrial baseline systems in terms of reconstruction SNR. Future directions may include the problem formulation with other penalty terms in order to provide alternate ways to control the contents of the masked TF regions.

6. REFERENCES

- [1] K. Gröchening, *Foundations of Time-Frequency Analysis*, Birkhäuser, Boston (MA), 2011.
- [2] 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.
- [3] E. Cano, D. FitzGerald, A. Liutkus, M. D. Plumbley, and F.-R. Stoter, “Musical source separation: An introduction,” *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 31–40, Jan. 2019.
- [4] F. Auger, P. Flandrin, Y.-T. Lin, S. McLaughlin, S. Meignen, T. Oberlin, and H.-T. Wu, “Time-frequency reassignment and synchrosqueezing: An overview,” *IEEE Signal Process. Mag.*, vol. 30, no. 6, pp. 32–41, 2013.
- [5] I. Daubechies, J. Lu, and H.-T. Wu, “Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool,” *Appl. and Comp. Harm. Anal.*, vol. 30, no. 1, pp. 243–261, 2011.
- [6] G. Thakur and H.-T. Wu, “Synchrosqueezing-based recovery of instantaneous frequency from nonuniform samples,” *SIAM J. Math. Anal.*, vol. 43, pp. 2078–2095, 2011.
- [7] P. Flandrin, “Time-frequency filtering based on spectrogram zeros,” *arxiv.org*, 2015.
- [8] J. Le Roux, H. Kameoka, N. Ono, A. de Cheveigné, and S. Sagayama, *Computational auditory induction as a missing-data model-fitting problem with Bregman divergence*, vol. 53, Speech Communication, May-June 2011.
- [9] P. Smaragdis, B. Raj, and M. Shashanka, “Missing data imputation for spectral audio signals,” *IEEE Int. Workshop Mach. Learn. Signal Process.*, 2009.
- [10] ANSYS VRXPERIENCE, “Sound website,” <https://www.ansys.com/fr-fr/products/systems/ansys-vrxperience/sound>.
- [11] 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.
- [12] Z. Prusa, P. Balazs, and P. Sondergaard, “A non-iterative method for reconstruction of phase from STFT magnitude,” in *IEEE/ACM Trans. Audio Speech Lang. Process.*, 2017, vol. 25, pp. 1154–1164.
- [13] P. Depalle, R. Kronland-Martinet, and B. Torr  sani, “Time-frequency multipliers for sound synthesis,” *Proc. SPIE*, vol. 6701, 2007.
- [14] M. D  rfler and B. Torr  sani, “On the time-frequency representation of operators and generalized gabor multiplier approximations,” *Journal of Fourier Analysis and Applications*, vol. 16, pp. 261–293, 2010.
- [15] M. D  rfler, *Gabor analysis adapted to music*, Ph.D. thesis, University of Vienna, Austria, 2002.
- [16] H. G. Feichtinger and K. Nowak, *A first survey of Gabor multipliers*, Advances in Gabor Analysis, 2002.