



HAL
open science

Local time-frequency fading

Ama Marina Kreme, Bruno Torr sani, Antoine Deleforge

► **To cite this version:**

Ama Marina Kreme, Bruno Torr sani, Antoine Deleforge. Local time-frequency fading. ICA 22 - International Congress on Acoustics 2022, Oct 2022, Gyeongju, South Korea. hal-03923596

HAL Id: hal-03923596

<https://hal.science/hal-03923596>

Submitted on 4 Jan 2023

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.

Local time-frequency fading

Marina KRÉME¹; Bruno TORRÉSANI²; Antoine DELEFORGE³

^{1,3}Université de Lorraine, CNRS, Inria, Loria, F-54000 Nancy, France

²Aix-Marseille Univ, CNRS, I2M, Marseille, France

ABSTRACT

For efficient analysis of non-stationary signals, such as radar, sonar, speech, music, or general audio, time-frequency (TF) representations are required. They allow performing non-stationary filtering, ie extracting non-stationary "components" in signals. However, ad-hoc procedures are frequently used, whose accuracy is difficult to assess and control. In this contribution, we are interested in such time-frequency filtering, which we define as follows: the attenuation of a given, small region Ω in the time-frequency plane. This problem, addressed in a recent publication, is referred to as time-frequency fading (TFF).

We address a local version of TFF, called LTFF, that only exploits a neighborhood $\bar{\Omega}$ of the time-frequency region of interest Ω , which is supposed to be much smaller than the full time-frequency domain. This problem will be referred to as local time-frequency fading. We propose a method for determining such a neighborhood $\bar{\Omega}$. The local TFF problem is then solved using a variational approach, that minimizes a suitable quadratic objective function. Thus, there is a closed-form solution involving Gabor multipliers.

Keywords: Time-Frequency Fading, Local time-frequency fading, Gabor transforms, Gabor multipliers

1 INTRODUCTION

Filtering is one of the fundamental techniques in signal processing. By filtering one often means linear time invariant (LTI) filtering, which has a simple characterization in time and/or frequency domains. We refer to [1] for an in-depth description, in view of audio applications. Sticking to linear filters, time varying (LTV) filtering is also fundamental, as most devices actually feature time variations. LTV filters form a much wider class than LTI filters, and may be represented in several different ways, for example as matrices as described in [1] for digital signals, or pseudo-differential operators in the mathematical literature [2] in the analog case. Among LTV filters, filters that can be efficiently represented in a joint time-frequency domain have received significant interest in the last two decades. These include filters whose spectral characteristics (frequency response) vary slowly as a function of time, or filters designed to enhance, attenuate or cancel out components which possess specific localization properties in joint time-frequency domain (e.g., chirps with prescribed or parameterized frequency modulation laws, transient oscillatory waves). Several formulations for non-stationary Wiener filtering have been proposed, that rely on time-frequency representations such as the spectrogram [3] or Wigner distribution [4, 5, 6]. These generally rely on pointwise multiplication in the time-frequency domain prior to inversion, very much in the spirit of classical LTI filtering. Filtering in the time-frequency domain has been used for a wide range of applications including decision, detection, or time-frequency segmentation.

Time-frequency filtering exploits the ability of some time-frequency representations to efficiently represent signals of interest. Limited to linear transforms, usual choices are Gabor transform [7] or Short Time Fourier Transform [8], on which we will focus in the present paper, or wavelet transform, constant Q transform or S transform, the choice often depending on the application domain.

In this paper, we discuss a local version of time-frequency fading (TFF) [9], which is a type of time-frequency filter design. Time-frequency fading is the process of restoring a signal that has been degraded by an additive disturbance, in the case where the latter is well localized in a specific region of the time-frequency domain. We previously investigated and proposed a rigorous formulation of TFF in our previous paper [9]. We

¹ ama-marina.kreme@inria.fr

² bruno.torresani@univ-amu.fr

³ antoine.deleforge@inria.fr

solved it and proposed several algorithms that make use of random projection techniques to enable numerical scaling. TFF, on the other hand, formulates the problem in the signal dimension. As a result, despite the fact that dimension reduction methods have been used to approximate the solution, TFF remains difficult to apply to long signals, such as those sampled at 44100 Hz. Here, we rely on the formulation proposed in and take advantage of the support reduction. In section 2 we briefly recall the time-frequency fading problem, and we introduce and formulate local TFF. In section 3, we illustrate LTFF on an example of signal mixing. Section 4 presents the ongoing work. An extended version of this work is under preparation.

2 GLOBAL AND LOCAL TIME-FREQUENCY FADING

2.1 TFF and notations

The Time-Frequency Fading (TFF) problem introduced in [9] may be seen as a time-frequency analogue of a stop-band filter design problem, where one tries to attenuate a set of prescribed regions $\Omega_p \subset \Lambda$ of the time-frequency plane Λ , with prescribed gains. This is achieved via a variational formulation, of the form

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathbb{C}^L} \|\mathcal{V}_{\mathbf{g}} \mathbf{x} - \mathcal{V}_{\mathbf{g}} \mathbf{z}\|_{\Omega}^2 + \sum_{p=1}^P \lambda_p \|\mathcal{V}_{\mathbf{g}} \mathbf{x}\|_{\Omega_p}^2, \quad (1)$$

where $\mathbf{z} \in \mathbb{C}^L$ is the input signal, $\mathcal{V}_{\mathbf{g}} : \mathbb{C}^L \rightarrow \Lambda$ is a linear time-frequency transform (see below for notations), $\Omega_p \subset \Lambda$, $p = 1, \dots, P$ are the regions of interest in the time-frequency plane Λ , and the $\lambda_p \in \mathbb{R}_+$ are parameters that control the tradeoff between the various terms in (1). $\|\cdot\|$ is the usual Frobenius norm on the time-frequency space Λ , and $\|\cdot\|_{\Omega}$ denotes its restriction to a given region Ω .

The quadratic nature of the objective function naturally leads to a linear system, whose closed-form solution reads (see [9] for details)

$$\mathbf{x}^* = \mathbf{z} - \left(1 + \sum_{p=1}^P (1 - \lambda_p) \mathfrak{M}_{\Omega_p} \right)^{-1} \sum_{p=1}^P \lambda_p \mathfrak{M}_{\Omega_p} \mathbf{z}, \quad (2)$$

which can involve high dimensional linear system for large signals. In [9], dimension reduction techniques were proposed to handle the problem and use a spectral approach that exploits time-frequency localization properties. However, the formulation of [9] involves a L -dimensional problem, which may involve regions of the time-frequency plane quite far away from Ω and do not significantly influence the solution.

The approach can be developed for any linear time-frequency transform, but in the present paper we will limit to the discrete Gabor transform (DGT) defined as follows (we refer to [9] for more details and notations, and to [10, 11] for numerical implementations we use in this work). Let a, b be two divisors of L , and set $M = L/b$ and $N = L/a$. The discrete time-frequency space is denoted by $\Lambda = \mathbb{Z}_M \times \mathbb{Z}_N$. For any $\mathbf{x} \in \mathbb{C}^L$, its DGT $\mathcal{V}_{\mathbf{g}} \in \ell^2(\Lambda) = \mathbb{C}^{M \times N}$ is defined by

$$\mathcal{V}_{\mathbf{g}} \mathbf{x}[m, n] = \langle \mathbf{x}, \mathbf{g}_{mn} \rangle = \sum_{\ell=0}^{L-1} \mathbf{x}[\ell] \bar{\mathbf{g}}[\ell - na] e^{-2i\pi m \ell / M}, \quad (m, n) \in \Lambda, \quad (3)$$

where the $\mathbf{g}_{mn} \in \mathbb{C}^L$, called Gabor atoms, are obtained from \mathbf{g} by time-frequency translations, i.e. $\mathbf{g}_{mn}[\ell] = e^{2i\pi m \ell / M} \mathbf{g}[\ell - na]$. For convenience we introduce the time-frequency shift operators π_{mn} such that $\mathbf{g}_{mn} = \pi_{mn} \mathbf{g}$. These operators satisfy the composition rule $\pi_{mn} \pi_{m'n'} = e^{-2i\pi m'n'a/M} \pi_{m+m', n+n'}$, and the adjoint operator reads $\pi_{mn}^* = e^{-2i\pi mna/M} \pi_{-m, -n}$.

Given $\mathbf{M} \in \mathbb{C}^{M \times N}$, the corresponding *Gabor multiplier* is the linear operator $\mathfrak{M}_{\mathbf{M}} : \mathbb{C}^L \rightarrow \mathbb{C}^L$ given by

$$\mathfrak{M}_{\mathbf{M}} = \mathcal{V}_{\mathbf{g}}^* \mathbf{M} \mathcal{V}_{\mathbf{g}}, \quad (4)$$

where by abuse of notation we have denoted by \mathbf{M} the operator of pointwise multiplication with \mathbf{M} . \mathbf{M} is called time-frequency mask, or upper symbol, of $\mathfrak{M}_{\mathbf{M}}$. When \mathbf{M} is the indicator function of a subset $\Omega \subset \Lambda$ we will use the notation \mathfrak{M}_{Ω} . Elementary properties of Gabor multipliers can be found in [12]. Among these, the approach of [9] relies strongly on the fact that for real valued masks, corresponding Gabor multipliers are self-adjoint, and their spectrum can be controlled under suitable assumptions on the window \mathbf{g} .

The closed form solution of the problem (1) can be expressed in terms of Gabor multipliers, and will appear as a special case of the solution given below, we thus refrain to reproduce it in this section.

2.2 Local TFF

In this section, we are interested in the estimation of a part of a locally degraded signal. We define an envelope around the masked region, and formulate LTFE from the newly defined regions.

2.2.1 Determination of the time-frequency regions of interest

Let $\varepsilon > 0$ be a fixed tolerance parameter. For each sub-region Ω_p , we define the *influence envelope* Ω_p^ε of Ω_p as follows:

$$\Omega_p^\varepsilon = \left\{ (m, n) \in \Lambda : \exists (m', n') \in \Omega_p, |\langle \mathbf{g}_{mn}, \mathbf{g}_{m'n'} \rangle|^2 \geq \varepsilon \right\} \quad (5)$$

The set Ω_p^ε consists in the time-frequency points where the corresponding atoms "communicate" with those in the masked region Ω_p , and therefore depends on the window \mathbf{g} and the sampling parameters. Notice that

$$\langle \mathbf{g}_{mn}, \mathbf{g}_{m'n'} \rangle = \langle \pi_{m'n'}^* \pi_{mn} \mathbf{g}, \mathbf{g} \rangle = e^{2i\pi(m-m')n'a/M} \langle \pi_{m-m'} \pi_{n-n'} \mathbf{g}, \mathbf{g} \rangle = e^{2i\pi(m-m')n'a/M} \mathcal{A}_{\mathbf{g}}[m-m', n-n'] , \quad (6)$$

$\mathcal{A}_{\mathbf{g}}$ being the ambiguity function of the window \mathbf{g} : $\mathcal{A}_{\mathbf{g}}[m, n] = \langle \pi_{mn} \mathbf{g}, \mathbf{g} \rangle$. Therefore the ambiguity function of the analysis window \mathbf{g} is used as the selection criterion. We also denote by

$$\Delta_p = \Omega_p^\varepsilon \setminus \Omega_p , \quad \Delta = \bigcup_p \Delta_p , \quad \Gamma = \overline{\Omega} \setminus \Delta \quad (7)$$

respectively the complementary region of Ω_p in Ω_p^ε , the union of these complementary regions and the complementary region of Δ in $\overline{\Omega}$, and $\overline{\Omega} = \Lambda \setminus (\bigcup_p \Omega_p)$.

We will also need the projection of the envelope Ω_p^ε onto the time domain, which is the interval obtained as follows:

$$I_p = [an_{\min}(p), an_{\max}(p)], \quad (8)$$

where $n_{\min}(p) = \min \{ n \mid \exists m, (m, n) \in \Omega_p^\varepsilon \}$ and $n_{\max}(p) = \max \{ n \mid \exists m, (m, n) \in \Omega_p^\varepsilon \}$ are respectively the minimum and maximum time indices in the time-frequency envelope Ω_p^ε (we recall that a is the time sampling step, i.e. the hop size).

2.2.2 Local TFF formulation

Similarly to TFF, and using the above notations, we formulate local TFF as follows:

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathbb{C}^L} \sum_{p=1}^P \|\mathcal{V}_{\mathbf{g}} \mathbf{x} - \mathcal{V}_{\mathbf{g}} \mathbf{z}\|_{\Delta_p}^2 + \sum_{p=1}^P \lambda_p \|\mathcal{V}_{\mathbf{g}} \mathbf{x}\|_{\Omega_p}^2 + \|\mathcal{V}_{\mathbf{g}} \mathbf{x} - \mathcal{V}_{\mathbf{g}} \mathbf{z}\|_{\Gamma}^2 , \quad (9)$$

In this formulation, we keep the part of the signal that is not masked (terms 1 and 3) and control the energy within each sub-region Ω_p (term 2) Note that, at this point, the reconstructed signal has length L . The following resolution will make the concept of "local" more clear. By setting the gradient of the objective function in (1) equal to zero, we obtain the following linear system:

$$\left[\mathfrak{M}_{\Gamma} + \sum_{p=1}^P (\mathfrak{M}_{\Delta_p} + \lambda_p \mathfrak{M}_{\Omega_p}) \right] \mathbf{x} = \left[\mathfrak{M}_{\Gamma} + \sum_{p=1}^P \mathfrak{M}_{\Delta_p} \right] \mathbf{z} , \quad (10)$$

which we now approximate as a family of smaller dimensional problems.

For all $q = 1, \dots, P$, denote by 1_q the indicator function of I_q . Assuming that the intervals I_q are disjoint $\bigcap_{q=1}^P I_q = \emptyset$, we have

$$\mathbf{x} = \sum_{q=1}^P 1_q \mathbf{x} + 1_J \mathbf{x}, \quad (11)$$

where 1_j the complementary of $\cup_{q=1}^P I_q$. The linear system becomes (10):

$$\left[\sum_p \sum_q (\mathfrak{M}_{\Delta_p} + \lambda_p \mathfrak{M}_{\Omega_p}) 1_q + \sum_q \mathfrak{M}_{\Gamma} 1_q \right] \mathbf{x} = \left[\sum_p \sum_q \mathfrak{M}_{\Omega_p} 1_q + \sum_q \mathfrak{M}_{\Gamma} 1_q \right] \mathbf{z}$$

Assuming that the non-diagonal terms in the sums can be neglected, i.e. that

$$\|(\mathfrak{M}_{\Delta_p} + \lambda_p \mathfrak{M}_{\Omega_p}) 1_q\|, \quad \|\mathfrak{M}_{\Omega_p} 1_q\|$$

are small enough for all $q \neq p$, and

$$\|\mathfrak{M}_{\Gamma} 1_q\|$$

are small enough for all q , we end up with

$$[\mathfrak{M}_{\Delta_p} + \lambda_p \mathfrak{M}_{\Omega_p}] \mathbf{x}_p = \mathfrak{M}_{\Omega_p} \mathbf{z}_p, \quad (12)$$

where we have set $\mathbf{x}_p = 1_p \mathbf{x}$. The operator $\mathfrak{M}_{\Delta_p} + \lambda_p \mathfrak{M}_{\Omega_p}$ is in fact a Gabor multiplier associated to a constant mask $\mathbf{M} = 1_{\Delta_p} + \lambda_p 1_{\Omega_p}$, where 1_{Δ_p} (respectively 1_{Ω_p}) is the indicator function of the subregion Ω_p (respectively Δ_p). If $\lambda_p > 0$ for all p , then the smallest value of the mask \mathbf{M} is strictly positive, so the mask is said to be semi-normalized¹, which implies invertibility (see Proposition 3.7 in [13]). From now on, we assume that $\lambda_p > 0$ for all p . The solution (12) is then written :

$$\mathbf{x}_p = [\mathfrak{M}_{\Delta_p} + \lambda_p \mathfrak{M}_{\Omega_p}]^{-1} \mathfrak{M}_{\Omega_p} \mathbf{z}_p \quad (13)$$

Thus, for each sub-region p , the masked time domain signal is reconstructed. In this formula we have reduced the dimension of the problem. From a Gabor multiplier, initially of dimension L , we pass to a multiplier in the dimension of the masked region which is much smaller.

3 Experiments

We illustrate the approach on a (single) synthetic mixture of two real audio signals. We first describe the signals, but we do not recall the actions of Gabor multipliers already widely discussed in [9]. We also consider the case where $P = 1$, i.e. a single masked region.

3.1 Experimental setting

We consider two audio signals sampled at 8 kHz of length 16384 (about 2 seconds). The signal of interest is a car engine noise and the perturbation signal is the song of a bird. As mentioned in the previous sections, the signal of interest has a broadband spectrogram (i.e., with energy spread all over the time-frequency plane), while the perturbation signal has its energy well localized in the time-frequency plane. The Gabor transform for each of these signals is calculated with a Gauss window of length 256, the time-frequency lattice parameters are set to $a = 32$ and $b = 512$, generating a 513×256 time-frequency matrix. The observation is a linear combinations of these two signals, as shown in Fig 1 (left). The corresponding binary mask is on the right. The goal is to filter the bird song from the mixture. We construct the Ω_p^ε region as described in section 2.2.1. The value of ε here is 10^{-6} . The Fig 2 shows the masked Ω_p region (in dark yellow) as well as the Ω_p^ε region (in light yellow). The corresponding time interval is $I_p = [47 \ 127]$.

We then restrict the frequency time plane of dimension of size 513×256 to 513×80 . We solve TFF from this reduced time-frequency plane. The time domain signal to be reconstructed corresponds only to the time domain signal which is in the Ω_p region and is of length 1600 here. The closed-form solution (12) depends on the regularization parameter $\lambda_p = \lambda > 0$ which is adjusted using the strategy presented in [14]: the optimal value of λ is chosen as the one for which the energy of the reconstructed signal in the region Ω_p corresponds to a given target energy E . Here, the energy E was chosen roughly by hand. In this example, the optimal value of λ is 10^{-3} .

¹A sequence $\mathbf{M} = (\mathbf{M}[n])_{n>1}$ is said to be semi-normalized if there exist a and b such that $0 < a \leq |\mathbf{M}[n]| \leq b < \infty$ for all $n \in \mathbb{N}$.

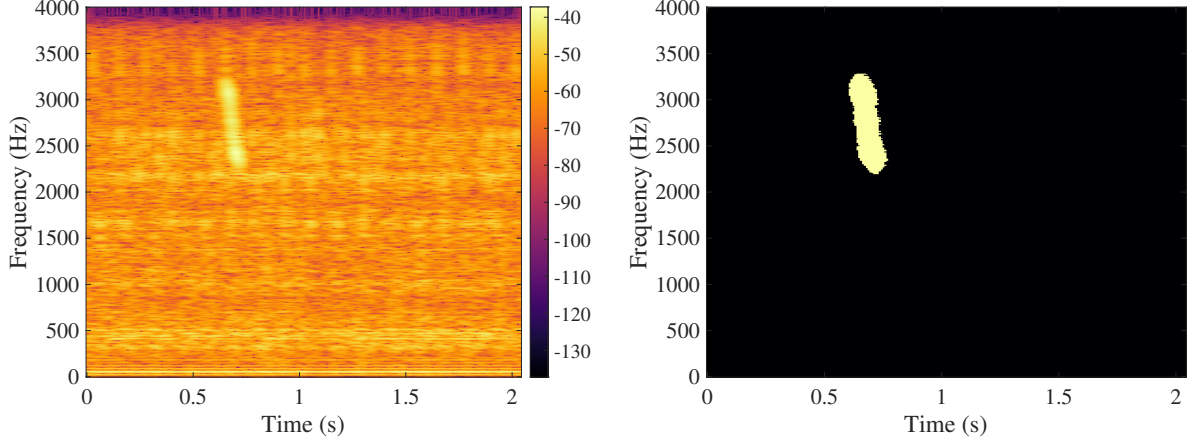


Figure 1. Left: Spectrogram of the mixture of the engine noise of a car (energy spread everywhere in the time-frequency plane) and of a bird "cuicui" (energy well localized - dark yellow area). Right: binary mask constructed from the spectrogram of two sources as in [9]

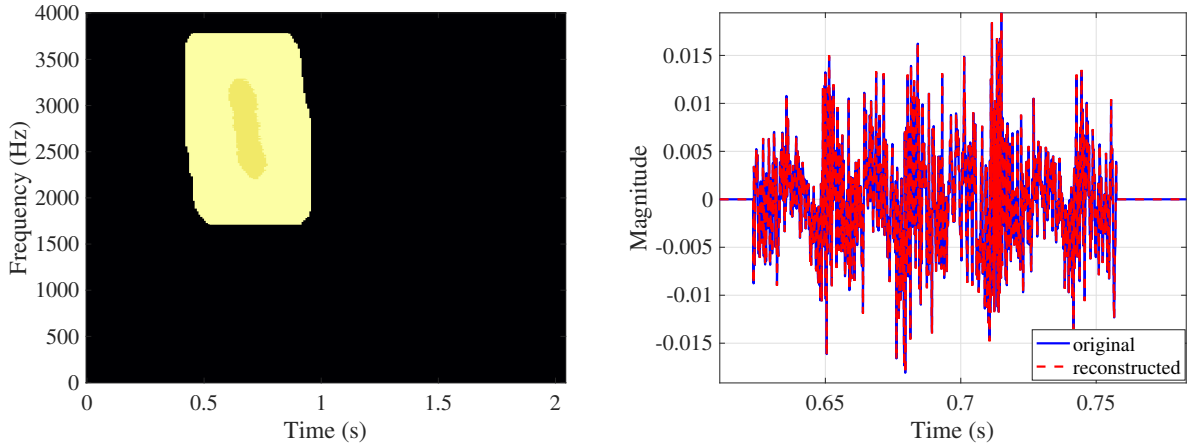


Figure 2. Left: Illustration of Ω_p and Ω_p^e regions in the case of a single region. The value of e is 10^{-6} . Right: a zoom on reconstruction of the part of the signal of interest masked by the bird song.

We quantitatively evaluate the quality of the reconstruction, via the source-to-distortion ratio (SDR) [15] expressed in decibel by

$$\text{SDR}(\mathbf{x}_p, \mathbf{x}_p^{ref}) = 20 \log_{10} \left(\frac{\|\mathbf{x}_p^{ref}\|}{\|\mathbf{x}_p - \mathbf{x}_p^{ref}\|} \right), \quad (14)$$

where \mathbf{x}_p^{ref} is the reference signal. The right panel of Figure 2 shows the reconstruction of the part of the car engine noise signal masked by the bird song in red. We can see that on this simple example, LTFF reconstructs almost perfectly the missing part of the signal with $\text{SDR} = 40$ dB. This experiment shows that LTFF could lead to satisfactory reconstruction results in the case of several regions $P > 1$. Moreover, LTFF could solve such problems in high dimension since it has fast computation time.

4 Ongoing work

We present a novel method for performing local time-frequency fading in this paper. Because it is limited to a neighborhood of the masked region, this method can handle longer signals. However, the obtained theoretical

results are based on assumptions concerning a control on the norm of the sum of the two Gabor multipliers that we have not presented here. As part of this work, we will establish more rigorous proof results. From an experimental point of view, we have presented one example to illustrate the idea of our work. In the rest of our work, we will extend these experiments to the case where we have $P > 1$ regions and also to signals sampled at 44.1 kHz. We will also compare the results of reconstructions obtained with TFF as well as the computation times. The experiments will then be extended to a larger class of signals by automating the code. Finally, the combination of TFF and LTFF may result in a novel source separation method.

References

- [1] J. O. Smith, *Introduction to Digital Filters: With Audio Applications*, ser. Music signal processing series. W3K, 2007. [Online]. Available: <https://books.google.fr/books?id=pCl1CQUAsHEC>
- [2] 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,” vol. 12, pp. 43–55, 2005.
- [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] G. F. Boudreaux-Bartels and T. W. Parks, “Time-varying filtering and signal estimation using Wigner distribution synthesis techniques,” vol. 34, no. 3, pp. 442–451, 1986.
- [5] F. Hlawatsch, G. Matz, H. Kirchauer, and W. Kozek, “Time-frequency formulation, design, and implementation of time-varying optimal filters for signal estimation,” vol. 48, no. 5, pp. 1417–1432, 2000.
- [6] L. Stankovic, “On the time-frequency analysis based filtering,” vol. 55, no. 5, pp. 216–225, May 2000. [Online]. Available: <https://doi.org/10.1007/BF02994785>
- [7] 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.
- [8] M. R. Portnoff, “Time-frequency representation of digital signals and systems based on short-time Fourier analysis,” vol. 28, no. 1, pp. 55–69, 1980.
- [9] A. M. Krémé, V. Emiya, C. Chaux, and B. Torrèsani, “Time-frequency fading algorithms based on Gabor multipliers,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 1, pp. 65–77, Jan 2021.
- [10] 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. 1 250 032–1 – 1 250 032–27, 2012.
- [11] Z. Průša, P. L. Søndergaard, N. Holighaus, C. Wiesmeyer, and P. Balazs, “The Large Time-Frequency Analysis Toolbox 2.0,” in *Sound, Music, and Motion*, ser. LNCS. Springer International Publishing, 2014, pp. 419–442.
- [12] H. G. Feichtinger and K. Nowak, “A first survey of Gabor multipliers,” in *Advances in Gabor Analysis*, H. G. Feichtinger and T. Strohmer, Eds. Birkhäuser, 2002, pp. 99–128.
- [13] 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. R. S. Casey, K. Okoudjou and B. Sadler, Eds. Springer, 2020, to appear, preprint arXiv:1803.00415.
- [14] A. M. Krémé, V. Emiya, C. Chaux, and B. Torrèsani, “Filtering out time-frequency areas using Gabor multipliers,” 2020, pp. 5790–5794.
- [15] E. Vincent, R. Gribonval, and C. Févotte, “Performance measurement in blind audio source separation,” vol. 14, no. 4, pp. 1462–1469, Jul. 2006.