PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Adhoc method to invert the reassigned time-frequency representation

Shristi RAJBAMSHI [1]; Peter BALAZS [2]; Nicki HOLIGHAUS [3]

Acoustics Research Institute, Austria

Abstract

In this contribution, we present a heuristic method to invert the reassigned short time Fourier transform magnitude spectrum to allow the reconstruction of the original time domain signal. This is a simple method just involving an additional smearing step before phase retrieval. Finally, we provide some numerical evidence that our method combined with existing phase retrieval methods shows excellent performance in synthesizing pure tones and synthetic signals from the reassigned coefficients. Furthermore, we also illustrate the practical importance of our results by applying our method to a real-world audio signal.

Keywords: Time-frequency representation, time-frequency reassignment, invertibility

1 INTRODUCTION

Time-Frequency (TF) representation techniques such as the Short Time Fourier Transform (STFT) (1, 3), particularly its magnitude, have been long used to analyze non-stationary signals. Despite being an efficient tool for analysis of complex signals, STFT often suffer from temporal-spectral resolution trade-off problem (2), thereby resulting in a smeared visual representation. In order to tackle this problem, the concept of TF *reassignment* was originated. This method, first introduced by Kodera (4) and later popularized by Auger and Flandrin (5), takes care of the aforementioned problem by utilizing the usually discarded phase information to produce a sharper TF representation. In spite of this clear advantage, TF reassignment has not received much attention because of its lack of invertibility, i.e., reconstruction of the original signal is not possible as the phase information is not available for synthesis. So far, there have been very few studies investigating the ways to reconstruct the original signal from the reassigned coefficients. One such method based on additive sound modeling can be read in (6) and (7). In this paper, we present an approach to invert the reassigned STFT magnitude spectrum by approximating the original STFT magnitude spectrum in combination with an existing phase retrieval method to allow the synthesis of the original signal.

1.1 Notation

In the following, we will denote matrices with uppercase letters, such as A, and column vectors as lowercase letters, such as a. $a[i,j]$ will denote an element of the matrix A where i and j denotes the row and column index respectively. We will write variables as lowercase letters, e.g., s and constants as uppercase letters, e.g., S.

For a matrix A, the *frobenius* norm is defined as $\|\mathrm{A}\|_{fro} = (\sum_i \sum_j |a[i,j]|^2)^{\frac{1}{2}}$. For a vector $\mathrm{x} = [x_0, x_1, \cdots, x_{L-1}]^\mathsf{T}$, the ℓ_1 norm is defined as $\|\mathrm{x}\|_1 = \sum_{l=0}^{L-1} |x_l|$.

[1] shristi.rajbamshi@oeaw.ac.at
[2] peter.balazs@oeaw.ac.at
[3] nicki.holighaus@oeaw.ac.at

2 OVERVIEW OF TIME-FREQUENCY REASSIGNMENT

As per (5), TF reassignment technique can be applied to various TF representations. However, for simplicity and clarity, we will only restrict to the STFT magnitude in this paper.

We will start by recalling a sub-sampled version of the discrete STFT called the Discrete Gabor Transform (DGT) (9). For any finite and real signal $\mathrm{x} \in \mathbb{R}^L$, the DGT with respect to the *analysis window* $\mathrm{w} \in \mathbb{R}^L$, time step $a \in \mathbb{N}$ and number of modulations $M \in \mathbb{N}$ is given by,

$$\mathrm{X_w}[n,m] = \sum_{l=0}^{L-1} \mathrm{x}[l+na]\mathrm{w}[l]e^{-i2\pi ml/M}, \tag{1}$$

where $n \in \{0,\dots,N-1\}$ denotes the time-shift index, $m \in \{0,\dots,M-1\}$ denotes the modulation index and [l+na] is assumed to be evaluated modulo L according to circular indexing. The transform *redundancy* is given by the ratio M/a.

If $a = 1$ and $M = L$, then (1) represents the full STFT, aka sliding FFT. In polar coordinate form, (1) can be decomposed as,

$$\mathrm{X_w}[n,m] = |\mathrm{X_w}[n,m]|e^{i\Phi[n,m]}, \tag{2}$$

where $|\mathrm{X_w}[n,m]|$ represents the magnitude and $\Phi[n,m]$ represents the phase.

Furthermore, the inverse DGT (IDGT) with respect to the *synthesis window* $\mathrm{w}_d \in \mathbb{R}$ is given as,

$$\bar{\mathrm{x}}[l] = \sum_{n=0}^{N-1}\sum_{m=0}^{M-1} \mathrm{X_w}[n,m]\mathrm{w}_d[l-na]e^{i2\pi m(l-na)/M}, \tag{3}$$

where $l \in \{0,\dots,L-1\}$. Note that, if w_d is a *canonical dual* window for w (9), then $\bar{\mathrm{x}}$ is equivalent to x up to some numerical precision error.

Independent of the choice of window in (1), the STFT always suffers from uncertainty (2) manifesting in a visual representation either smeared in time, in frequency or in both directions, that cannot be arbitrarily reduced. In order to refine this blurred representation, reassignment technique is utilized. In general, reassignment maps the TF data to the "true region of TF support" of the analyzed signal by means of the partial derivatives of the phase of the underlying TF representation. These partial derivatives are called instantaneous time-frequency estimates. In our case, reassignment maps the magnitude at position (n,m) to a different TF position $(\tilde{n},\tilde{m})$ based on the phase derivatives of the underlying STFT. These new TF positions are referred to as reassignment operators and are computed as following (5, 10),

$$\tilde{n}[n,m] = n + \mathrm{Re}\left[\frac{\mathrm{X}_{\frac{1}{T_{SR}}t\mathrm{w}}[n,m]}{\mathrm{X_w}[n,m]}\right], \tag{4}$$

$$\tilde{m}[n,m] = m - \frac{L}{2\pi}\mathrm{Im}\left[\frac{\mathrm{X}_{T_{SR}d\mathrm{w}}[n,m]}{\mathrm{X_w}[n,m]}\right], \tag{5}$$

where, $\mathrm{X_w}$, $\mathrm{X}_{\frac{1}{T_{SR}}t\mathrm{w}}$ and $\mathrm{X}_{T_{SR}d\mathrm{w}}$ are the DGTs of the signal x corresponding to the analysis window w, time-weighted version of the window w, i.e., $\left(\frac{1}{T_{SR}}t\mathrm{w}\right)[l] = \frac{1}{T_{SR}}t \cdot w(t)|_{t=lT_{SR}}$ and derivative of the windows with respect to time, i.e., $(T_{SR}d\mathrm{w})[l] = T_{SR}\frac{dw(t)}{dt}|_{t=lT_{SR}}$. Note that, T_{SR} is the sampling period and we consider samples of a continuous window function $w(t)$, i.e. $\mathrm{w}[l] = w(t)|_{t=lT_{SR}}$.

Hence, the reassigned TFR for any point (n',m') is then calculated as (5),

$$\mathrm{R}^{\mathrm{x}}_{\mathrm{w}}[n',m'] = \sum_{m}^{M-1}\sum_{n}^{-N-1} \mathrm{S_w}[n,m]\delta(n'-\tilde{n}[n,m])\delta(m'-\tilde{m}[n,m]). \tag{6}$$

where $\mathrm{S_w}[n,m] = |\mathrm{X_w}[n,m]|$.

3 PROPOSED METHOD TO OBTAIN THE ORIGINAL TIME-FREQUENCY REPRESENTATION

Although TF reassignment has many desirable properties such as TF shift invariance, energy conservation, perfect localization for pure sinusoids, chirps and impulses (5), etc, the major drawback of this method is that it is not invertible which constrains its use. In order to address this problem, we propose a method to approximate the original TF representation from the reassigned TF representation by implementing a two step post processing technique: The first step involves convolution of the reassigned coefficients with a kernel and the second step involves recovery of the original STFT coefficients by employing existing phaseless retrieval methods.

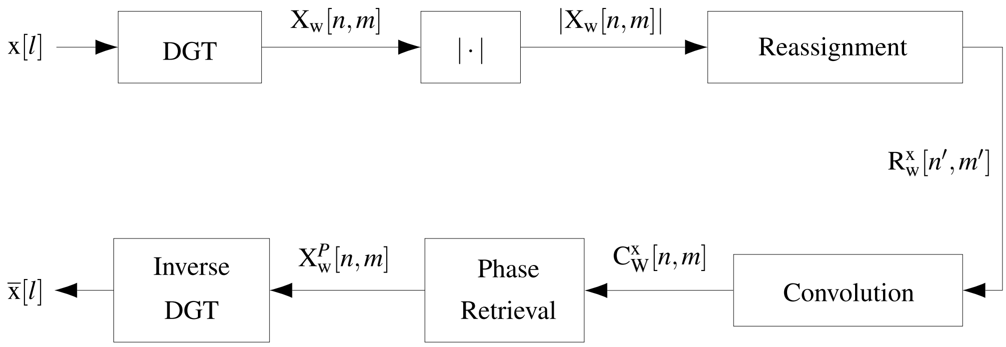


Figure 1. Block diagram of the proposed inverse reassignment technique

3.1 Convolution as Inverse Reassignment

In order to recover the original STFT magnitude coefficients from the reassigned coefficients $\mathrm{R}^{\mathrm{x}}_{\mathrm{w}}[n',m']$, we perform two-dimensional convolution of the reassigned coefficients with a kernel. We will refer this procedure as *inverse reassignment*. Mathematically, it can be formulated as,

$$\mathrm{C}^{\mathrm{x}}_{\mathrm{W}}[n,m] = (\mathrm{R}^{\mathrm{x}}_{\mathrm{w}} * \mathrm{W})[n,m] = \sum_{m'}\sum_{n'} \mathrm{R}^{\mathrm{x}}_{\mathrm{w}}[n',m']\mathrm{W}[n-n',m-m'] \tag{7}$$

where $\mathrm{W}[n,m]$ is the convolution kernel. In our work, we choose $\mathrm{W}[n,m] = \mathrm{w}[n] \otimes \hat{\mathrm{w}}[m]$, where w is the same *window* with respect to which the STFT was computed and $\hat{\mathrm{w}}$ is its discrete Fourier transform given by,

$$\hat{\mathrm{w}}[m] = \frac{1}{\sqrt{K}} \sum_{k=0}^{K-1} \mathrm{w}[k] e^{-i\frac{2\pi}{K}mk} \tag{8}$$

where $m \in \{0,\dots,M-1\}$. Further, we make an assumption that both w and $\hat{\mathrm{w}}$ are symmetric, localized around 0 and $\|\mathrm{w}\|_1 = \|\hat{\mathrm{w}}\|_1 = 1$.

Proof in case of complex exponential

We now present a proof for the feasibility of our approach for the particular case of complex exponential incorporating the concepts discussed in the previous sections. For that, we define a discrete complex exponential signal $\mathrm{x} \in \mathbb{R}^L$ of frequency f_1, i.e., $\mathrm{x}[l] = Ae^{i\frac{2\pi}{L}f_1 l}$. Using (1), and choosing $a = 1$ and $M = L$, the STFT of x is,

$$\mathrm{X}_{\mathrm{w}}[n,m] = Ae^{i\frac{2\pi}{L}f_1 n}\hat{\mathrm{w}}[m-f_1], \tag{9}$$

the magnitude is,

$$\mathrm{S}_{\mathrm{w}}[n,m] = \mid \mathrm{X}_{\mathrm{w}}[n,m] \mid = A\hat{\mathrm{w}}[m-f_1], \tag{10}$$

and the reassigned coefficients are,

$$R_{\mathrm{w}}^{\mathrm{x}}[n',m'] = \begin{cases} A & \text{if } m' = f_1 \\ 0 & \text{otherwise} \end{cases}.$$

With the convolution kernel, $\mathrm{W}[n,m] = \mathrm{w}[n] \otimes \hat{\mathrm{w}}[m]$, *inverse reassignment* yields,

$$\begin{aligned} \mathrm{C}_{\mathrm{w}}^{\mathrm{x}}[n,m] &= (\mathrm{R}_{\mathrm{w}}^{\mathrm{x}} * \mathrm{W})[n,m] \\ &= \sum_{m'}\sum_{n'} \mathrm{R}_{\mathrm{w}}^{\mathrm{x}}[n',m']\mathrm{W}[n-n',m-m'] \\ &= \sum_{m'} \hat{\mathrm{w}}[m-m'] \sum_{n'} \mathrm{R}_{\mathrm{w}}^{\mathrm{x}}[n',m']\mathrm{w}[n-n'] \end{aligned} \tag{11}$$

Now, if $m' = f_1$,

$$\begin{aligned} \mathrm{C}_{\mathrm{w}}^{\mathrm{x}}[n,m] &= \hat{\mathrm{w}}[m-f_1] \sum_{n'} \mathrm{R}_{\mathrm{w}}^{\mathrm{x}}[n',f_1]\mathrm{w}[n-n'] \\ &= \hat{\mathrm{w}}[m-f_1] \sum_{n'} A\mathrm{w}[n-n'] \\ &= A\hat{\mathrm{w}}[m-f_1]\underbrace{\sum_{n'} \mathrm{w}[n-n']}_{=1} \qquad \because \|\mathrm{w}\|_1 = 1 \\ &= A\hat{\mathrm{w}}[m-f_1]. \end{aligned} \tag{12}$$

Comparing (10) and (12), we obtain,

$$\boxed{\mathrm{C}_{\mathrm{w}}^{\mathrm{x}}[n,m] = |\mathrm{X}_w(n,m)|}. \tag{13}$$

Hence, we can easily prove for the particular case of complex exponential that the operation of inverse reassignment yields the original magnitude.

3.2 Phase Retrieval

The recovered STFT magnitude, i.e., $\mathrm{C}_{\mathrm{w}}^{\mathrm{x}}[n,m]$, is then used to estimate the original STFT coefficients employing a Phase Retrieval (PR) method. PR can be defined as a method to recover the original signal from the magnitude only measurements. There are several studies focused on phase retrieval. In this section, we give a summary of three specific phase retrieval methods that we use in our work to synthesize the signal from the reassigned coefficients. We will denote them as PR-I, PR-II and PR-III. The first PR method is based on an algorithm called Phase Gradient Heap Integration (PGHI) (12), second is based on an algorithm called Griffin-Lim (GL) (11, 16) and third is based on the algorthim called Fast Griffin-Lim (13).

In PGHI, phase is retrieved by using the gradient theorem based on the assumption that the phase gradient and phase at one point is known. Phase gradient are obtained from the magnitude gradient. Since this a non-iterative method, the algorithm is very fast and appropriate for long signals.

The GL algorithm is an iterative approach which converges to the original phase by repeatedly computing the STFTs and inverse STFTs. First it performs an inverse STFT using the original magnitude and then computes the STFT of the obtained times series. Next, it combines the original magnitude with the phase of the newly computed STFT and then again the inverse STFT is calculated. These steps are repeated till the algorithm converges. This algorithm is not very efficient as it converges very slowly.

The FGL algorithm is a phase retrieval method that treats the phase recovery problem as an optimization problem. As the name suggests, this algorithm is a faster version of the traditional GL algorithm.

4 SIMULATION RESULT

To evaluate the performance of our proposed method, we carry out experiments using four different test signals. All the simulations are implemented in MATLAB using the LTFAT toolbox (14, 15). Our first test signal is a *complex exponential* of 50 Hz generated as $x[l] = e^{i2\pi 50l}$, second is a *synthetic signal* generated as $x[l] = cos(4800\pi l + 350cos(10\pi l)) + cos(20000\pi l - 5400cos(2\pi(l-1))) + cos(750\pi l + 80cos(4\pi(l-0.03)))$, third is the (in)famous audio signal called "glockenspiel" (14) and fourth is the speech signal called "linus" (14).

Details about all the test signals, such as their lengths (L), sampling frequencies (fs) and parameters, a and M required for computing corresponding DGTs and IDGTs are enlisted in Table 1. For all the signals, the DGT is computed with respect the sampled and periodized Gaussian window with variance 1 and IDGT is computed with respect the canonical dual of the window used for DGT.

Table 1

	L (in samples)	fs (in Hz)	a	M
Complex Exponential	400	400	1	400
Synthetic	44100	44100	200	1000
Glockenspiel	262144	44100	200	2000
Linus	41461	8000	200	1000

For each simulation, DGT, reassignment, convolution and IDGT are computed using 'dgt','gabreassign', 'pconv' and 'idgt' fucntions available in the LTFAT toolbox. Additionally, the phase retrieval is carried out using 'constructPhase' and 'frsynabs' routines which are also available in the LTFAT toolbox. Note that the routines 'constructPhase' correspond to PR-I and 'frsynabs' with parameters *griflim* and *fgriflim* correspond to PR-II and PR-III respectively.

4.1 Evaluation Metric

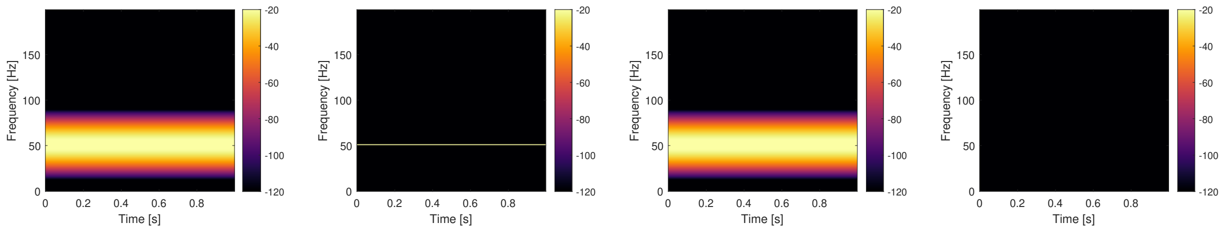
For each signal, we will reconstruct the original signal by employing various phase retrieval methods discussed in Section 3.2 and calculate the consequent spectral convergence by including and excluding the convolution step discussed in Section 3.1. We will use original phase as the benchmark to compare the performance of various PR methods. The spectral convergence is given as (16),

$$20\log_{10}\left(\frac{\|X-\bar{X}\|_{fro}}{\|X\|_{fro}}\right).$$

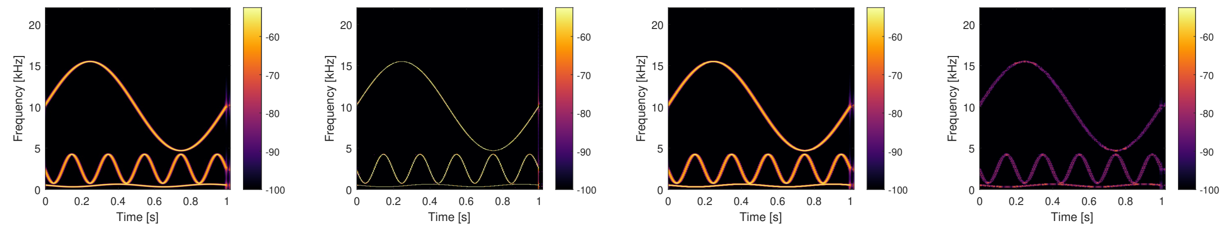
where $\bar{X}$ is the reconstructed STFT computed from the re-synthesized signal $\bar{x}$ using 1.

4.2 Discussion

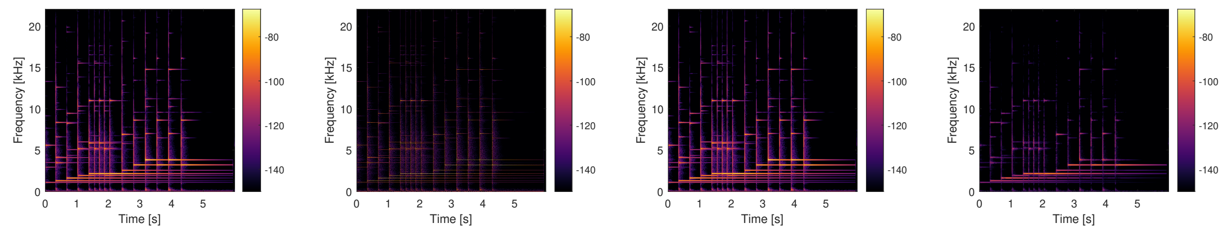
In Figure 2(a), it can be seen that our proposed method works perfectly for a pure tone as the difference between the original and the reconstructed STFT magnitude is zero (up to numerical precision). Also for the synthetic signal (see 2(b)), the reconstruction error is very small. Similar performance can be observed for more complex signals like glockenspiel and linus (see 2(c) and 2(d)). It should be noted that our method performs well not only for the case of high redundancy, for example, in case of complex exponential when the redundancy is 400, but also for the low redundancy case, for example in case of other three signals when the redundancy is 5, 10 and 5 respectively. Remember that, redundancy is calculated as M/a.



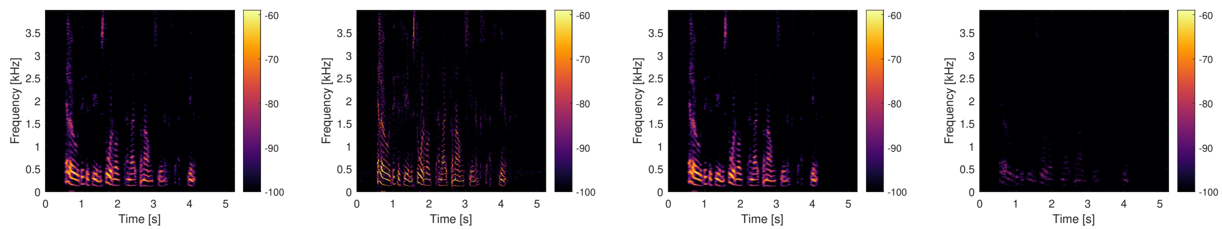
(a) Complex Exponential



(b) Synthetic Signal



(c) Glockenspiel



(d) Linus

(I) (II) (III) (IV)

Figure 2. (I) *left*: Original STFT magnitude, (II) *left middle*: Reassigned STFT magnitude, (III) *right middle*: Reconstructed STFT magnitude (Inverse Reassignment), (IV) *right*: difference between (I) and (III)

The Table 2 depicts the spectral convergence, using different phase retrieval methods while including and excluding the convolution step denoted as "C" and "NC" respectively.

Table 2. Spectral Convergence

	Complex Exponential		Synthetic		Glockenspiel		Linus	
	NC	C	NC	C	NC	C	NC	C
Original Phase	-7.6 dB	-169.63 dB	-8.67 dB	-28.1 dB	-8.79 dB	-31.06 dB	-9.27 dB	-25.7 dB
PR-I	-7.65 dB	-169.63 dB	-7.89 dB	-13.02 dB	-6.64 dB	-14.95 dB	-8.46 dB	-10.3 dB
PR-II	-7.65 dB	-21.19 dB	-8.04 dB	-16.62 dB	-6.72 dB	-14.39 dB	-8.69 dB	-16.55 dB
PR-III	-7.65 dB	-23.86 dB	-8.10 dB	-19.20 dB	-6.78 dB	-14.73 dB	-8.70 dB	-17.46 dB

It can be deduced from Table 2 there is significant improvement in spectral convergence if we convolve the reassigned coefficient with the kernel before we perform the phase retrieval.

5 CONCLUSION

We have demonstrated that reassigned STFT magnitude can be inverted using the convolution technique based on the assumption that the analysis window and its Fourier transform are symmetric and their ℓ_1-norm is 1. The proposed inverse reassignment method in combination with the existing PR techniques can be used to recover the original time domain signal. Our method exhibited excellent performance not only for pure tones but also for complex signals like audio and speech. Furthermore, pilot tests indicated that there was no audible difference between the original and reconstructed audio and speech signals. Therefore, we believe that this method can be very useful in extending the usage of reassignment technique for present and future signal processing applications which has been limited thus far due to the lack of invertibility.

ACKNOWLEDGEMENT

This work was supported by the Austrian Science Fund (FWF) START-project FLAME ("Frames and Linear Operators for Acoustical Modeling and Parameter Estimation"; Y 551-N13).

REFERENCES

1. Gröchenig K. Foundations of Time-Frequency Analysis, Applied Numerical Harmonic Analysis, Birkhäuser, Boston, MA, 2001.

2. Gröcheni, K. Uncertainty Principles for Time-Frequency Representations, Chapter 2 in Advances in Gabor Analysis (eds. H. G. Feichtinger and T. Strohmer), Birkhäuser Boston, 2003, pp 11-30.

3. Cohen, L. Time-Frequency Analysis: Theory and Applications, Prentice Hall Signal Processing Series, Prentice Hall, 1995

4. Kodera, K.; Gendrin R.; Villedary C. Analysis of time-varying signals with small bt values, IEEE Transactions on Acoustics, Speech and Signal Processing, Vol 26, Feb 1978, pp 64–76.

5. Auger F.; Flandrin P. Improving the readability of time-frequency and time-scale representation by the reassignment method, IEEE Transactions on Signal Processing, Vol 43(5), May 1995, pp 1068–1088.

6. Fitz, K.; Haken, L. On the use of time frequency reassignment in additive sound modeling, Journal of the Audio Engineering Society, Vol 50(11), Nov 2002, pp 879-93.

7. Han, L., Sacchi, MD. Spectral decomposition and de-noising via time-frequency and space-wavenumber reassignment, Geophysical Prospecting, Vol 62(2), Mar 2014, pp 244-57.

8. Søndergaard, P.L. Gabor frames by Sampling and Periodization, Advanced in Computational Mathematics, Vol. 27(4), Nov 2007, pp 355 –373.

9. Søndergaard, P.L.; Hansen P.C.; Christensen, O. Finite discrete Gabor analysis, Ph.D. dissertation, Technical University of Denmark, 2007.

10. Fenet, S.; Badeau, R.; Richard, G. Reassigned time–frequency representations of discrete time signals and application to the Constant-Q Transform, Signal Processing, Elsevier,Vol 132, Mar 2017, pp 170-6.

11. Griffin, D.; Lim, J. Signal estimation from modified short-time Fourier transform, IEEE Transactions on Acoustics, Speech and Signal Processing, Vol 32(2), 1984, pp 236-243.

12. Prusa, Z.; Balazs, P.; Søndergaard P.L. A non-iterative method for STFT phase (re)Construction, IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol 25(5), May 2017, pp 1154-64.

13. Perraudin, N.; Balazs, P.; Søndergaard, P.L. A fast Griffin-Lim algorithm, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, October 2013, pp 1-4.

14. Søndergaard, P.L.; Torresani, B.; Balazs, P. The linear time frequency analysis toolbox, International Journal of Wavelets, Multiresolution and Information Processing, Vol 10(04), 2012, p. 1250032.

15. Prusa, Z.; Søndergaard, P.L., Holighaus, N.; Wiesmeyr, C.; Balazs, P. The Large Time-Frequency Analysis Toolbox 2.0 in Sound, Music, and Motion, ser. LNCS. Springer International Publishing, 2014, pp 419–442.

16. Sturmel, N.; Daudet, L. Signal reconstruction from STFT magnitude: A state of the art. International conference on digital audio effects (DAFx), Sep 2011, pp 375-386.