Toward High-Quality Real-Time Signal Reconstruction From STFT Magnitude

Zdeněk Průša and Pavel Rajmic

***Abstract*—An efficient algorithm for real-time signal reconstruction from the magnitude of the short-time Fourier transform (STFT) is introduced. The proposed approach combines the strengths of two previously published algorithms: the real-time phase gradient heap integration and the Gnann and Spiertz's real-time iterative spectrogram inversion with look-ahead. An extensive comparison with the state-of-the-art algorithms in a reproducible manner is presented.**

***Index Terms*—Phase reconstruction, real-time, short-time Fourier transform (STFT), spectrogram, time–frequency.**

I. Introduction

In TIME–FREQUENCY signal processing, it is a common practice to work only with the magnitude of the short-time Fourier transform (STFT) of a signal. However, as soon as reconstruction is desired, phase information becomes essential. When the magnitude is modified, it is often sufficient to reuse the original phase to recover the signal [1]; however, some spectrogram modifications might invalidate the phase and the reconstruction procedure can therefore lead to undesired artifacts [2]. In some applications, the original phase is not available at all [3]. STFT phase retrieval algorithms alleviate these problems by allowing complete disposal of the existing phase and constructing a new valid phase from scratch, taking the modified magnitude. Unfortunately, currently available STFT phase retrieval algorithms cannot always be expected to fulfil all possible requirements at the same time. For example, some algorithms require the knowledge of the entire magnitude component and they typically need a large number of costly iterations to produce a good result [4]–[7]. This fact disqualifies them from being used in any real-time or interactive applications. Algorithms capable of processing signals in real-time, i.e., in the frame-by-frame manner with bounded delay [8]–[11], tend to produce noticeable artifacts such as "phasiness" [2], metallic ringing, and echo for specific classes of audio signals.

Manuscript received December 15, 2016; revised March 20, 2017; accepted April 7, 2017. Date of publication April 25, 2017; date of current version May 4, 2017. This paper was supported in part by the Austrian Science Fund (FWF): (Y 551–N13), in part by the joint project of the FWF and the Czech Science Foundation (GAČR) numbers I 3067-N30 and 17-33798L, respectively, and in part by the National Sustainability Program under Grant LO1401. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Balázs Bank. *(Corresponding author: Zdeněk Průša.)*

Z. Průša is with the Acoustics Research Institute, Austrian Academy of Sciences, Vienna 1040, Austria (e-mail: zdenek.prusa@oeaw.ac.at).

P. Rajmic is with the Signal Processing Laboratory, Faculty of Electrical Engineering and Communication, Brno University of Technology, Brno 61600, Czech Republic (e-mail: rajmic@feec.vutbr.cz).

Digital Object Identifier 10.1109/LSP.2017.2696970

In this letter, we propose a real-time phase reconstruction algorithm which outperforms the state-of-the-art algorithms by a large margin with respect to both the objective performance evaluation and the perceived quality of the reconstruction. We compare our method with the following algorithms, which can be considered as state-of-the-art algorithms: the real-time iterative spectrogram inversion (RTISI) [12] later improved by including look-ahead frames [8], [13] (RTISI-LA), and further modified slightly in line with the work of Gnann and Spiertz [14]–[16] and Gnann [17] (GSRTISI-LA). From the point of view of this letter, the crucial property of GSRTISI-LA is that it allows defining an initial phase estimate of the latest look-ahead frame.

In our previous work [18], we have proposed a noniterative algorithm termed real-time phase gradient heap integration (RTPGHI). RTPGHI is based on the phase-magnitude relationship, which allows estimating the phase increments between neighboring STFT coefficients solely from the magnitude. The algorithm requires one look-ahead frame (zero look-ahead frame version is also available) and, as it turns out, it is a suitable candidate for providing the initial phase guess for GSRTISI-LA. In this letter, we combine the good transient behavior of RTPGHI with the outstanding properties of GSRTISI-LA to perform a high-quality signal reconstruction from a spectrogram. We aim at a high reconstruction quality, and therefore, we do not include noniterative algorithms presented in [9]–[11] in our comparison. Although they are much faster than iterative algorithms, they produce results of significantly lower quality. The only exception is the single pass spectrogram inversion algorithm [9] (SPSI), which we included in the evaluation as an alternative way of phase initialization.

In the spirit of reproducible research, the implementation of the algorithms, audio examples as well as the scripts reproducing the experiments are available at http://ltfat.github.io/notes/048. The code depends on our MATLAB/GNU Octave [19] packages LTFAT [20], [21] (version 2.1.3 or above) and PHASERET (version 0.2.0 or above). Both toolboxes are open source and they can be obtained from http://ltfat.github.io and http://ltfat.github.io/phaseret , respectively.

II. STFT and Its Inverse

The discrete STFT of an input signal $f \in \ell^2(\mathbb{Z})$ using the analysis window $g \in \ell^2(\mathbb{Z})$ is defined as

$$c_n(m) = \left(\mathcal{V}_g f\right)_n (m) = \sum_{l\in\mathbb{Z}} f(l+na)\overline{g(l)}\mathrm{e}^{-\mathrm{i}2\pi m l/M} \qquad (1)$$

where the overline denotes the complex conjugation, M is a finite number of frequency channels indexed with $m = 0, \ldots, M-1$, $n \in \mathbb{Z}$ is the time-frame index, and the parameter a acts as the time step (window shift) in samples. The window g will further be considered to be real, whole-point symmetric, and compactly supported such that the range of summation can be reduced to $\mathcal{I} = \{-\lfloor \mathrm{len}\,(g)/2 \rfloor, \ldots, \lceil \mathrm{len}\,(g)/2 \rceil - 1\}$, where $\mathrm{len}\,(g)$ is the length of the window support.

The synthesis window $\widetilde{g}$ can be obtained as

$$\widetilde{g}(l) = \frac{1}{M} \frac{g(l)}{\sum_{n\in\mathbb{Z}} g(l-na)^2} \tag{2}$$

if the following two conditions are met: The window support length g is less or equal to the number of frequency channels, i.e., $\mathrm{len}\,(g) \leq M$, and there is a nonempty overlap between windows, i.e., $\mathrm{len}\,(g) > a$. Under these assumptions, the sum in the denominator in (2) is nonzero and a-periodic, $\widetilde{g}$ and g have identical time support and the following relation holds:

$$\sum_{n\in\mathbb{Z}} g(l-na)\widetilde{g}(l-na) \equiv 1/M. \tag{3}$$

Please refer to [22]–[24], for example, for a thorough mathematical treatment of the invertibility of the discrete STFT (also referred to as the Discrete Gabor transform) in the context of the frame theory. The Gabor frame theory calls the window computed using (2) the *canonical dual window* and the inequality $\mathrm{len}\,(g) \leq M$ is usually referred to as the *painless condition* [25].

Having the synthesis window $\widetilde{g}$, the individual time-frames f_n of f can be recovered from the respective coefficients using

$$f_n(l) = \begin{cases} \widetilde{g}(l) \sum_{m=0}^{M-1} c_n(m) \mathrm{e}^{\mathrm{i}2\pi m l/M} & \text{for } l \in \mathcal{I}, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Consequently, a partial signal reconstruction from time frames up to the index N is given by

$$\widetilde{f}_N(l) = \sum_{n=-\infty}^{N} f_n(l-na) \tag{5}$$

(cf., overlap-add procedure) and, clearly, the original signal is formally obtained by taking $N = \infty$.

III. Algorithms

In the following, we will denote the magnitude of the coefficients of the nth time-frame as $s_n(m) = |c_n(m)|$. The goal of phase reconstruction algorithms is to estimate the unknown phase of the coefficients. Let us denote the estimated phase of $c_n(m)$ as $\widetilde{\phi}_n(m)$ and the estimated coefficient as $\widetilde{c}_n(m) = s_n(m)\mathrm{e}^{\mathrm{i}\widetilde{\phi}_n(m)}$. In real time, a particular time frame can thus be recovered by plugging coefficients $\widetilde{c}_n(m)$ into (4).

A. Overview of RTPGHI

The RTPGHI algorithm [18] is a real-time capable version of the PGHI algorithm [26]. It is an efficient, noniterative algorithm which, by itself, provides results of a good quality. In particular, in contrast to other algorithms, it does not introduce transient "smearing." Therefore, one expects RTPGHI to be a suitable candidate for initializing GSRTISI-LA.

The RTPGHI algorithm is based on the relationship between the gradients of the phase and the logarithm of the magnitude of STFT. It employs an adaptive integration scheme to recover the phase. The best performance is achieved using the Gaussian window, but other windows can be used as well. The RTPGHI algorithm comes in two versions, RTPGHI(1) requiring one look-ahead frame and RTPGHI(0) requiring no look-ahead frame. For details see the above mentioned references. Here, we only give a conceptual introduction to RTPGHI(1).

Let us denote $s_{\log,n}(m) = \log(s_n(m))$. The estimate of the scaled phase derivative in the frequency direction $\widetilde{\phi}_{\omega,n}(m)$ and in the time direction $\widetilde{\phi}_{t,n}(m)$ expressed solely using the magnitude can be written as

$$\widetilde{\phi}_{\omega,n}(m) = -\frac{\gamma}{2aM}\left(s_{\log,n+1}(m) - s_{\log,n-1}(m)\right)$$

$$\widetilde{\phi}_{t,n}(m) = \frac{aM}{2\gamma}\left(s_{\log,n}(m+1) - s_{\log,n}(m-1)\right) + \frac{2\pi am}{M}$$

with γ being the "width" parameter of the Gaussian window [26]. Given the phase estimate $\widetilde{\phi}_{n-1}(m)$, the phase $\widetilde{\phi}_n(m)$ for a particular m is computed using one of the following equations:

$$\widetilde{\phi}_n(m) \leftarrow \widetilde{\phi}_{n-1}(m) + \frac{1}{2}\left(\widetilde{\phi}_{t,n-1}(m) + \widetilde{\phi}_{t,n}(m)\right) \tag{6}$$

$$\widetilde{\phi}_n(m) \leftarrow \widetilde{\phi}_n(m-1) + \frac{1}{2}\left(\widetilde{\phi}_{\omega,n}(m-1) + \widetilde{\phi}_{\omega,n}(m)\right) \tag{7}$$

$$\widetilde{\phi}_n(m) \leftarrow \widetilde{\phi}_n(m+1) - \frac{1}{2}\left(\widetilde{\phi}_{\omega,n}(m+1) + \widetilde{\phi}_{\omega,n}(m)\right). \tag{8}$$

B. GSRTISI-LA With RTPGHI Initialization

In this section, we present a variant of GSRTISI-LA that enables using an arbitrary analysis window and allows free choice of the window overlap length and the number of frequency channels (as long as the conditions presented in Section II hold). As already mentioned, this proposed approach employs RTPGHI to find the initial phase estimate.

Assuming RTPGHI(1) is used for initialization, the algorithm processes one time frame at a time, taking into account N_{LA} future frames. The $N_{\mathrm{LA}} - 1$ look-ahead frames are used for the basic version of the GSRTISI-LA algorithm, one additional look-ahead frame is required for RTPGHI. In addition to specifying the windows g and $\widetilde{g}$, the algorithm requires N_{LA} additional analysis windows g_p, $p = 0, \ldots, N_{\mathrm{LA}} - 1$, which are obtained as

$$g_p(l) = M\frac{g(l)}{g_{\mathrm{sum}}(l+pa)}, \tag{9}$$

where

$$g_{\mathrm{sum}}(l) = \sum_{q=-\infty}^{N_{\mathrm{LA}}-1} g(l-qa)\widetilde{g}(l-qa). \tag{10}$$

The notation has been simplified in the formal description of the proposed algorithm RTPGHI(1) + GSRTISI-LA($N_{\mathrm{LA}} - 1$)

Algorithm 1: RTPGHI(1) + GSRTISI-LA($N_{\mathrm{LA}}-1$), $n-$th time frame.

Input: Number of look-ahead frames N_{LA}, number of iterations I, magnitude of STFT coefficients $s_n, \ldots, s_{n+N_{\mathrm{LA}}}$

Output: Time frame f_n.

Compute $f_{n+N_{\mathrm{LA}}-1}$ using (5) and coefficients $\widetilde{c}_{n+N_{\mathrm{LA}}-1}$ estimated using the RTPGHI algorithm (requires $s_{n+N_{\mathrm{LA}}}$)
for $i = 1, 2, \ldots, I$ **do**
  for $p = N_{\mathrm{LA}} - 1, \ldots, 0$ **do**
    Compute $\widetilde{f}_{n+N_{\mathrm{LA}}-1}$ using (6)
    $t \leftarrow \left(\mathcal{V}_{g_p} \widetilde{f}_{n+N_{\mathrm{LA}}-1}\right)_{n+p}$
    $c_{n+p} \leftarrow s_{n+p}\, t / |t|$
    Compute f_{n+p} using (5)
  end
end

in Algorithm 1. The indices in the brackets referring to the vector entries have been omitted; it is assumed that entire vectors are employed. The extension to RTPGHI(0) + GSRTISI-LA(N_{LA}) is straightforward.

Even though we operate with infinite sum limit, in practice, due to the finite support of the windows, at most $N_{\mathrm{LB}} = \lceil \mathrm{len}\,(g)\,/a \rceil - 1$ "look-back" frames is sufficient.

Please note that other types of phase initialization are possible. The authors of the original version of GSRTISI-LA proposed to perform simple *phase unwrapping* [16]. Another option is to employ algorithms such as SPSI [9] in place of RTPGHI. However, in our experience neither of these two approaches brings considerable improvements over the zero or random phase initialization. Often, using the described means of phase initialization is even harmful for the overall performance.

C. Real-Time Deadline, Delay and Computational Complexity

The worst-case execution time for a single output frame must be less than a/f_{s} seconds (time frame shift divided by the sampling rate in hertz) to meet the real-time deadline restriction. This fact limits the number of iterations I that can be performed, nevertheless the actual ceiling for I is entirely dependent on the computing power of the device. Since the number of look-ahead frames can be varied, we will further use the number of *per frame* iterations to be able to directly compare different settings.

The typical delay of the real-time STFT analysis-synthesis scheme is equal to the length of the window. Each look-ahead frame of the phase reconstruction algorithm increases the delay by the window shift a; therefore, the overall input–output delay is $(\mathrm{len}(g) + aN_{\mathrm{LA}})\,/f_{\mathrm{s}}$ seconds.

IV. Experiments

In the experiments, we used the SQAM database [27], which consists of 70 recordings sampled at 44.1 kHz. The first ten seconds from the first channel of each sound sample were used in the evaluation. For the performance comparison, we used the normalized mean-squared error between the original STFT magnitude s and the STFT magnitude of the reconstructed signal $\widetilde{f}$, previously referred to as *spectral convergence* [28]

$$\mathcal{C} = \sqrt{\frac{\sum_{n=0,m=0}^{N-1,M-1} \left(s_n(m) - \left|\left(\mathcal{V}_g \widetilde{f}\right)_n(m)\right|\right)^2}{\sum_{n=0,m=0}^{N-1,M-1} s_n(m)^2}} \tag{11}$$

where N denotes the total number of time frames of the finite signal. The transform $\mathcal{V}_g$ uses the same g, a and M as the transform used to obtain s. Values in decibels are obtained by computing $20\log_{10}\mathcal{C}$. Although not without shortcomings (see [28] for details), spectral convergence seems to be the only suitable error measure for evaluating phase reconstruction algorithms due to its robustness with respect to phase error irregularity. Due to the irregular behavior of the reconstructed phase, the recovered waveform is usually very different from the original signal and the energy of the time domain error signal is actually comparable to the energy of the signal itself. Therefore, the common time domain error measures such as the signal-to-noise ratio are not applicable. Similarly, in our experience, automated quality evaluation methods such as PEAQ [29] or PEMO-Q [30] are not suitable either. They mostly seem to be "blind" to subtle perceptual errors.

In turns out that a substantial window overlap is necessary in order to produce the results of high perceptual quality. In our tests, we use 87.5% window overlap, which results from using the time step size $a = 256$ together with the fixed number of frequency channels $M = 2048$, and the Gaussian window as the analysis window truncated at 1% of its height such that $\mathrm{len}\,(g) = 2048$. Using an even higher window overlap further improves the results.

Please note that whenever we refer to the average error in decibels, we mean $20\log_{10}\frac{1}{70}\sum_{k=1}^{70}\mathcal{C}_k$, where $\mathcal{C}_k$ is the error of the kth sound excerpt obtained from (11). Averaging errors that have been already converted to decibels (which is occasionally done in other contributions) produces even better (lower) errors for all the algorithms.

In the real-time setting, there is room only for a limited number of iterations, but since the exact number is device dependent, we will evaluate the performance of the algorithms for up to 200 per frame iterations.

Fig. 1 shows the average spectral convergence depending on the number of iterations and the number of look-ahead frames $N_{\mathrm{LA}} \in \{0, 1, 2, 3, 7\}$. Actually, $N_{\mathrm{LA}} = 7$ is the maximum number of look-ahead frames which directly overlap with the currently processed frame in our setup. Error measures obtained by the noniterative RTPGHI algorithm are indicated as horizontal dashed lines. Note that we are intentionally using a fixed range of values on the vertical axis. The values of the parameters (g, a and M) were chosen the same as that used in [26, Fig. 6(a)] to allow a direct comparison. Note that the proposed algorithm with $N_{\mathrm{LA}} = 7$ (corresponding to 87 ms input–output delay) outperforms even the best of the offline-only algorithms in terms of $\mathcal{C}$.

One can observe that the common behavior of the iterative algorithms is that the average spectral convergence initially decreases rapidly and from a certain number of

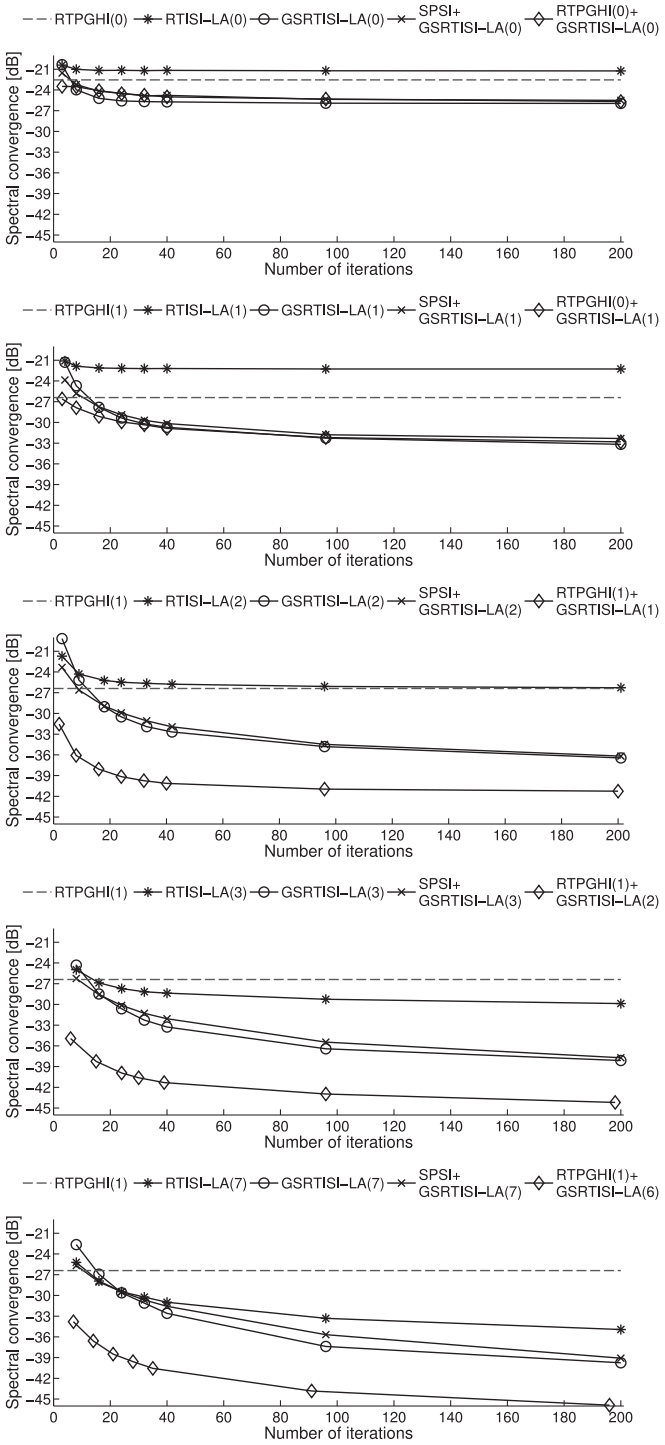

Fig. 1. Comparison of algorithms.

iterations upwards it starts to level off. This phenomenon could be explained by the fact that some signals in the database reach "convergence" at some point while others continue to improve. From Fig. 1, it is also clear that SPSI [9] initialization does not bring any significant improvement to the GSRTISI-LA algorithm. Furthermore, one can observe that the proposed algorithm clearly outperforms the others whenever two or more look-ahead frames are used. The scores for individual files for $N_{\mathrm{LA}} = 2$ ($N_{\mathrm{LA}} = 1$ in case of RTPGHI) and $I = 24$ per frame

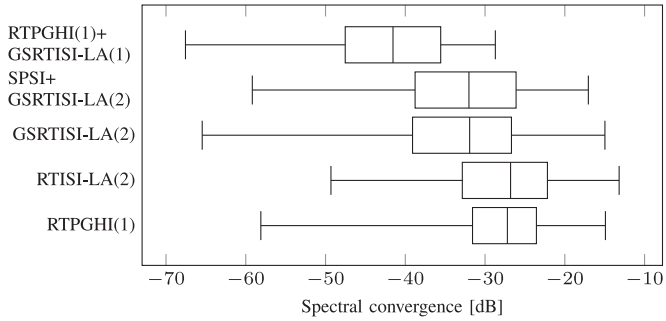

Fig. 2. Box plot (minimum, first quartile, median, third quartile, maximum) of the errors obtained for $N_{\mathrm{LA}} = 2$ ($N_{\mathrm{LA}} = 1$ in the case of RTPGHI) from the 70 sound excerpts.

iterations, as well as the sound examples for all the samples from the SQAM database, can be found at the accompanying web page http://ltfat.github.io/notes/048. Additionally, a box plot of the results is depicted in Fig. 2.

When inspecting the results obtained for individual sound excerpts, one can notice that the iterative algorithms struggle with reconstructing recordings of percussion instruments such as claves and castanets and with attacks of transients in general. Conveniently, the RTPGHI algorithm performs very well in such cases and the combination with GSRTISI-LA inherits and even improves upon the behavior as indicated by the significantly low maximum error in Fig. 2.

A real-time demo allowing one-to-one comparison of the algorithms is available in the PHASERET toolbox, implemented in `demo_blockproc_phaseret2.m`.

V. Conclusion

It has been shown that the combination of GSRTISI-LA and RTPGHI outperforms other algorithms and their combinations, as soon as enough look-ahead frames are used.

Although we have only presented objective error measures in this letter, in our experience, the quality of the reconstructed signal reflects the error measure improvement. An interested reader can verify this claim by listening to the sound samples found at the accompanying webpage or by running `demo_blockproc_phaseret2.m` using his/her custom audio examples.

In this letter, we have assumed that the phase is completely unknown and only the original clean magnitude is known. The proposed algorithm can be easily modified to respect and use coefficients with known phase, but, in the real-world, noisy or modified magnitudes and phases are usually observed. Therefore, as the future work, we will focus on simultaneous magnitude and phase estimation given corrupted, noisy or incomplete information since the phase-aware signal processing is currently an active field of research [31]–[33].

Acknowledgment

The authors thank T. Necciari and N. Holighaus for their valuable comments. The authors also thank the anonymous reviewers for their suggestions. Infrastructure of the SIX Center was used.

REFERENCES

[1] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 3, pp. 1066–1074, Mar. 2007.

[2] J. Laroche and M. Dolson, "Phase-vocoder: About this phasiness business," in *Proc. 1997 IEEE ASSP Workshop , Appl. Signal Process. Audio Acoust.*, Oct. 1997, pp. 1–4.

[3] P. Smaragdis, B. Raj, and M. Shashanka, "Missing data imputation for time-frequency representations of audio signals," *J. Signal Process. Syst.*, vol. 65, no. 3, pp. 361–370, 2011.

[4] D. Griffin and J. Lim, "Signal estimation from modified short-time Fourier transform," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 2, pp. 236–243, Apr. 1984.

[5] N. Perraudin, P. Balazs, and P. L. Søndergaard, "A fast Griffin-Lim algorithm," in *IEEE Workshop, Appl. Signal Process. Audio Acoust.*, Oct. 2013, pp. 1–4.

[6] R. Decorsiere, P. L. Søndergaard, E. N. MacDonald, and T. Dau, "Inversion of auditory spectrograms, traditional spectrograms, and other envelope representations," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 1, pp. 46–56, Jan. 2015.

[7] J. L. Roux, H. Kameoka, N. Ono, and S. Sagayama, "Fast signal reconstruction from magnitude STFT spectrogram based on spectrogram consistency," in *Proc. 13th Int. Conf. Digit. Audio Effects*, Sep. 2010, pp. 397–403.

[8] X. Zhu, G. T. Beauregard, and L. Wyse, "Real-time signal estimation from modified short-time Fourier transform magnitude spectra," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 5, pp. 1645–1653, Jul. 2007.

[9] G. T. Beauregard, M. Harish, and L. Wyse, "Single pass spectrogram inversion," in *Proc. IEEE Int. Conf., Digit. Signal Process.*, Jul. 2015, pp. 427–431.

[10] P. Margon, R. Badeau, and B. David, "Phase reconstruction of spectrograms with linear unwrapping: Application to audio signal restoration," in *Proc. 23rd Eur. Signal Process. Conf.*, Aug. 2015, pp. 1–5.

[11] M. Chami, J. D. Martino, L. Pierron, and E. H. Ibn-Elhaj, "Real-time signal reconstruction from short-time Fourier transform magnitude spectra using FPGAs," in *Proc. 5th. Int. Conf. Inf. Syst. Econ. Intell.*, Djerba, Tunisia, Feb. 2012.

[12] G. T. Beauregard, X. Zhu, and L. Wyse, "An efficient algorithm for real-time spectrogram inversion," in *Proc. 8th Int. Conf. Digit. Audio Effects*, Sep. 2005, pp. 1–6.

[13] X. Zhu, G. T. Beauregard, and L. Wyse, "Real-time iterative spectrum inversion with look-ahead," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2006, pp. 229–232.

[14] V. Gnann and M. Spiertz, "Comb-filter free audio mixing using STFT magnitude spectra and phase estimation," in *Proc. 11th Int. Conf. Digit. Audio Effects*, Sep. 2008, pp. 357–361.

[15] V. Gnann and M. Spiertz, "Inversion of STFT magnitude spectrograms with adaptive window lengths," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2009, pp. 325–328.

[16] V. Gnann and M. Spiertz, "Improving RTISI phase estimation with energy order and phase unwrapping," in *Proc. 13th Int. Conf. Digit. Audio Effects*, Sep. 2010, pp. 1–5.

[17] V. Gnann, *Signal Reconstruction from Multiresolution Magnitude Spectrograms for Audio Signal Processing* (ser. Aachen Series on Multimedia and Communications Engineering). Herzogenrath, Germany: Shaker Verlag, Feb. 2014, vol. 13.

[18] Z. Průša and P. L. Søndergaard, "Real-Time Spectrogram Inversion Using Phase Gradient Heap Integration," in *Proc. Int. Conf. Digit. Audio Effects*, Sep. 2016, pp. 17–21.

[19] J. W. Eaton, D. Bateman, S. Hauberg, and R. Wehbring, *GNU Octave version 4.0.0 manual: A High-Level Interactive Language for Numerical Computations*, 2015. [Online]. Available: http://www.gnu.org/software/octave/doc/interpreter

[20] P. L. Søndergaard, B. Torrésani, and P. Balazs, "The Linear Time Frequency Analysis Toolbox," *Int. J. Wavelets, Multiresolution Anal. Inf. Process.*, vol. 10, no. 4, pp. 1–27, 2012.

[21] Z. Průša, P. L. Søndergaard, N. Holighaus, C. Wiesmeyr, and P. Balazs, "The Large Time-Frequency Analysis Toolbox 2.0," in *Sound, Music, and Motion* (ser. Lecture Notes in Computer Science). New York, NY, USA: Springer, 2014, pp. 419–442.

[22] T. Strohmer, *Numerical Algorithms for Discrete Gabor Expansions*. Cambridge, MA, USA: Birkhäuser Boston, 1998, ch. 8, pp. 267–294.

[23] P. L. Søndergaard, "Finite discrete Gabor analysis," Ph.D. dissertation, Tech. Univ. Denmark, Kongens Lyngby, Denmark, 2007. [Online]. Available from: http://ltfat.github.io/notes/ltfatnote003.pdf

[24] P. L. Søndergaard, "Efficient algorithms for the discrete Gabor transform with a long FIR window," *J. Fourier Anal. Appl.*, vol. 18, no. 3, pp. 456–470, 2012.

[25] I. Daubechies, A. Grossmann, and Y. Meyer, "Painless nonorthogonal expansions," *J. Math. Phys.*, vol. 27, no. 5, pp. 1271–1283, 1986.

[26] Z. Průša, P. Balazs, and P. L. Søndergaard, "A non-iterative method for reconstruction of phase from STFT magnitude," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 5, pp. 1154–1164, May 2017.

[27] "Tech 3253: Sound Quality Assessment Material recordings for subjective tests," The European Broadcasting Union, Geneva, Tech. Rep., Sep. 2008. [Online]. Available: https://tech.ebu.ch/docs/tech/tech3253.pdf

[28] N. Sturmel and L. Daudet, "Signal reconstruction from STFT magnitude: A state of the art," in *Proc. 14th Int. Conf. Digit. Audio Effects*, pp. 375–386, 2011.

[29] T. Thiede *et al.*, "PEAQ—The ITU standard for objective measurement of perceived audio quality," *J. Audio Eng. Soc.*, vol. 48, no. 1-2, pp. 3–29, 2000. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=12078

[30] R. Huber and B. Kollmeier, "PEMO-Q: A new method for objective audio quality assessment using a model of auditory perception," *IEEE Trans., Audio, Speech, Lang. Process.*, vol. 14, no. 6, pp. 1902–1911, Nov. 2006.

[31] T. Gerkmann, M. Krawczyk-Becker, and J. L. Roux, "Phase processing for single-channel speech enhancement: History and recent advances," *IEEE, Signal Process. Mag.*, vol. 32, no. 2, pp. 55–66, Mar. 2015.

[32] P. Mowlaee, J. Kulmer, J. Stahl, and F. Mayer, *Single Channel Phase-Aware Signal Processing in Speech Communication: Theory and Practice*. Hoboken, NJ, USA: Wiley, 2016.

[33] P. Mowlaee, R. Saeidi, and Y. Stylianou, "Advances in phase-aware signal processing in speech communication," *Speech Commun.*, vol. 81, pp. 1–29, 2016.