PHASE CORRECTED TOTAL VARIATION FOR AUDIO SIGNALS

Kohei Yatabe and Yasuhiro Oikawa

Department of Intermedia Art and Science, Waseda University, Tokyo, Japan

ABSTRACT

In optimization-based signal processing, the so-called prior term models the desired signal, and therefore its design is the key factor to achieve a good performance. For audio signals, the time-directional total variation applied to a spectrogram in combination with phase correction has been proposed recently to model sinusoidal components of the signal. Although it is a promising prior, its applicability might be restricted to some extent because of the mismatch of the assumption to the signal. In this paper, based upon the previously proposed one, an improved prior for audio signals named instantaneous phase corrected total variation (iPCTV) is proposed. It can handle wider range of audio signals owing to the instantaneous phase correction term calculated from the observed signal.

Index Terms— Spectrogram, phase-aware processing, phase derivative, instantaneous frequency, convex optimization.

1. INTRODUCTION

For optimization-based signal processing methods, design of the so-called prior term, which imposes the prior knowledge about the desired signal, is important to achieve a good performance. While a complicated model is useful to obtain higher performance, a simple model is also important in practice. Recent methods may utilize several priors simultaneously to impose multiple aspects of the signal, and thus each prior is preferred to be simple as possible for reducing the overall complexity.

As such a simple prior term for audio signals, phase corrected total variation (PCTV) has been introduced in [1]. Total variation (composition of the first order difference and the ℓ_1-norm) is a quite popular prior, especially in image processing [2–5], that induces piece-wise smoothness to the signal. Based on the observation that a purely sinusoidal signal is represented by a smooth complex spectrogram after a suitable phase correction, PCTV is defined as the total variation applied time-directionally to the phase corrected spectrogram (see Section 2.1). Since it utilizes information of the phase spectrogram explicitly, PCTV can be regarded as a phase-aware prior, where such phase-aware methods receive much attention recently [6, 7]. It can easily be optimized by convex optimization techniques [8–12] because of its simplicity, and therefore PCTV is a promising prior which should be investigated further.

Although PCTV has many attractive properties, its applicability might be restricted to some extent because of the mismatch of the assumption to the signal. The key factor of PCTV is the phase correction which realizes the smooth time-frequency representation of a sinusoidal signal. However, the conventional PCTV deals with the phase in terms of the center frequency of each bin of the spectrogram. That is, the phase correction is performed not based on the phase of the signal but based on the parameters of a time-frequency analysis method. This mismatch might require a highly redundant time-frequency representation which may not be suitable for many applications of acoustical signal processing because of the computational complexity caused by the high redundancy.

In this paper, an improved PCTV, namely instantaneous phase corrected total variation (iPCTV), is proposed. It corrects the spectrogram based on the instantaneous phase of the signal so that the sinusoidal components are handled more appropriately. As the result of considering instantaneous phase, iPCTV can be applied to a broader range of spectrograms calculated by a larger shifting step of the window function than the conventional PCTV. For demonstrating its performance, a simple denoising problem is considered and is solved by the primal-dual splitting algorithm.

2. PHASE CORRECTED TOTAL VARIATION

In this section, after briefly reviewing the concept of the conventional PCTV, instantaneous correction of the phase based on the signal is proposed to improve the performance of PCTV.

2.1. The simple prior (Conventional PCTV) [1]

Let the short-time Fourier transform (STFT), or discrete Gabor transform, of a signal x with a window function w be defined as [13–15]

$$(\mathscr{F}^w x)(m,n) = \sum_{l=0}^{L-1} x(l+an)\overline{w(l)e^{2\pi ibml/L}}, \tag{1}$$

where $\overline{z}$ is complex conjugate of z, $i = \sqrt{-1}$, n and m are the time and frequency indices, and a and b are the time and frequency shifting steps, respectively. Since a sinusoidal signal (with initial phase ϕ_0) can be written as

$$s = e^{2\pi i(bfan/L+\phi_0)} = e^{2\pi i(bfa(n+1)/L+\phi_0)}e^{-2\pi ibfa/L}, \tag{2}$$

its STFT has the neighborhood relation,

$$(\mathscr{F}^w s)(f,n+1)e^{-2\pi ibfa/L} = (\mathscr{F}^w s)(f,n), \tag{3}$$

when f coincides with some m. From this equation, it can be seen that the time-directional difference of the adjacent components of STFT is zero for a sinusoidal component of $f = m$ when the phase factor $e^{-2\pi ibfa/L}$ is multiplied.

In [1], the phase corrected version of STFT is considered,

$$(\mathscr{F}^w_{\mathrm{PC}} x)(m,n) = (\mathscr{F}^w x)(m,n)e^{-2\pi ibman/L}, \tag{4}$$

which can be directly written as another form of STFT:

$$(\mathscr{F}^w_{\mathrm{PC}} x)(m,n) = \sum_{l=0}^{L-1} x(l)\overline{w(l-an)e^{2\pi ibml/L}}. \tag{5}$$

Then, PCTV was defined as time-directional total variation of the phase corrected STFT[1]:

$$\mathrm{TV}_{\mathrm{PC}}(x) = \|D_t \mathscr{F}^w_{\mathrm{PC}} x\|_1 = \|D_t E_{\mathrm{PC}} \mathscr{F}^w x\|_1, \tag{6}$$

[1]Note that this definition of PCTV is a reinterpreted version of the original description, where its motivation and description are detailed in [1].

where $\|\cdot\|_p$ is the p-norm, D_t is the time-directional difference $(D_t z)(m,n) = z(m,n) - z(m,n-1)$, and $(E_{\mathrm{PC}} z)(m,n) = z(m,n)e^{-2\pi i b m a n/L}$. This prior yields a small value when the signal is composed of sinusoidal components because of the relation in Eq. (3). On the other hand, non-sinusoidal components, such as random noise, are penalized. Many important audio signals, including speech and music, consist of sinusoidal components, and therefore PCTV is useful to model them. Note that PCTV induces peace-wise smoothness not only to the magnitude of the spectrogram but also to the phase, i.e., PCTV is a phase-aware prior [6, 7].

However, as pointed out in the original paper [1], PCTV might not work as expected when the frequency of the sinusoidal component f does not coincide with the center frequency of any subband of the spectrogram. In that case, $\mathrm{TV}_{\mathrm{PC}}(s)$ may not yield a small value because of the mismatch of the corrected phase. Although the amount of this mismatch can be reduced by increasing the number of subbands, it comes with a price of the computational cost which spoil the advantage of simplicity. Therefore, it is necessary to reduce the mismatch without increasing the computational cost much.

2.2. Instantaneous phase correction of STFT

In order to reduce the frequency mismatch, in addition to the constant correction in Eq. (4), instantaneous phase correction of the phase spectrogram derived from the signal is proposed here.

A sinusoidal signal at the m-th subband can be written as

$$\begin{aligned}\tilde{s} &= e^{2\pi i b(m+\delta)an/L} \\ &= e^{2\pi i b(m+\delta)a(n+1)/L} e^{-2\pi i b m a/L} e^{-2\pi i b \delta a/L},\end{aligned} \tag{7}$$

where $\delta(m,n)$ is the amount of the mismatch of the frequency on the scale of m (for simplicity, the initial phase ϕ_0 is omitted as it does not affect the result). This equation indicates the following neighborhood relation of STFT:

$$(\mathscr{F}^w \tilde{s})(m,n+1)e^{-2\pi i b m a/L} e^{-2\pi i b \delta a/L} = (\mathscr{F}^w \tilde{s})(m,n), \tag{8}$$

which is also represented as

$$(\mathscr{F}^w_{\mathrm{PC}} \tilde{s})(m,n+1)e^{-2\pi i b \delta a/L} = (\mathscr{F}^w_{\mathrm{PC}} \tilde{s})(m,n). \tag{9}$$

Since this relationship holds for the sinusoidal signal of any frequency in contrast to Eq. (3), the phase of every subband can be corrected without the frequency mismatch, if δ is known.

For estimating δ from the signal, one simple choice is the reassignment method. Reassignment is the methodology of enhancing a time-frequency representation by calculating the partial derivatives of the phase of each time-frequency bin [16–22]. The desired quantity for the correction, or instantaneous frequency, is obtained as time-differential of the phase, which can be calculated numerically as [18, 21]

$$\delta(m,n) = -\frac{1}{b}\,\mathrm{Im}\left[\frac{(\mathscr{F}^{w'}_{\mathrm{PC}} x)(m,n)}{(\mathscr{F}^{w}_{\mathrm{PC}} x)(m,n)}\right], \tag{10}$$

where $w' = \partial w/\partial t$ is the time-derivative of the window function [22], and $\mathrm{Im}[z]$ is the imaginary part of z. Note that, for numerical stability, $\delta(m,n)$ corresponding to extremely small $|(\mathscr{F}^w_{\mathrm{PC}} x)(m,n)|$ may need special treatment because it may be infinity or a random number. Once δ is calculated from the two spectrograms, the neighborhood relation of STFT can be fully corrected.

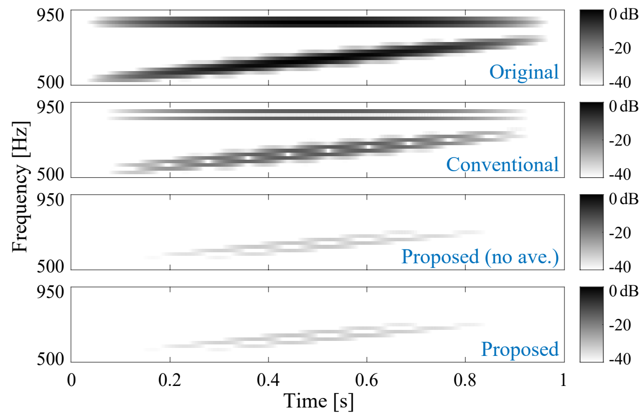


Fig. 1. Magnitude of the phase corrected difference (window: 2048, shift: 128 samples). The test signal, which consist of two sinusoidal components, were similar to that of the original paper of the conventional PCTV [1]. The third figure corresponds to iPCTV without the averaging [Eq. (13)], while the forth figure is iPCTV with the averaging [Eq. (12)].

2.3. Proposed prior: iPCTV

Based on the above observations, we propose iPCTV for audio signals which is defined as

$$\mathrm{TV}_{\mathrm{iPC}}(x) = \|D_t E_{\mathrm{iPC}} \mathscr{F}^w_{\mathrm{PC}} x\|_1 = \|D_t E_{\mathrm{iPC}} E_{\mathrm{PC}} \mathscr{F}^w x\|_1, \tag{11}$$

where $(E_{\mathrm{iPC}} z)(m,n) = z(m,n)e^{-2\pi i b a \tilde{\delta}(m,n)/L}$, and

$$\tilde{\delta}(m,n) = \begin{cases} \displaystyle\sum_{l=0}^{n-1} \frac{\delta(m,l+1)+\delta(m,l)}{2}, & (n \geq 1), \\ 0, & (n = 0), \end{cases} \tag{12}$$

is the averaged instantaneous phase [cumulative sum of the averaged instantaneous frequencies obtained by Eq. (10)]. This averaging is not necessary, but it slightly improves the effect of the correction as shown below (see Fig. 2). Note that the instantaneous phase factor $e^{-2\pi i b a \tilde{\delta}(m,n)/L}$ is calculated by Eq. (10) only once. That is, E_{iPC} is defined as element-wise multiplication of the constant factors which are calculated at first and regarded as constant thereafter. This definition of E_{iPC} is necessary to keep $\mathrm{TV}_{\mathrm{iPC}}$ convex as the original PCTV.

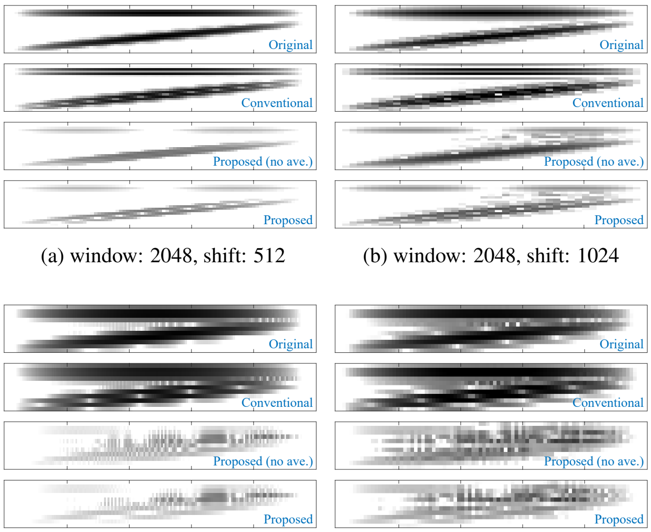
2.4. Some examples showing properties of iPCTV

To illustrate the characteristics of the proposed prior, several examples are shown. In addition to the conventional PCTV and the proposed iPCTV, a variant of iPCTV, which does not take the average of the phase as

$$\tilde{\delta}(m,n) = \sum_{l=0}^{n} \delta(m,l), \tag{13}$$

is also shown to demonstrate the effect of the averaging in Eq. (12). The test signal was composed of two sinusoidal components multiplied by the Hann window. The sampling frequency and the window function for STFT were, respectively, 44100 Hz and the Hann window.

Fig. 1 shows an example of the phase corrected difference of the test signal. The window length and the shifting step were 2048 and 128 samples, which were chosen for easier comparison to the



(a) window: 2048, shift: 512 (b) window: 2048, shift: 1024

(c) window: 1024, shift: 256 (d) window: 1024, shift: 512

Fig. 2. Magnitude of the phase corrected difference corresponding to several conditions of the window function. The top row is for the Hann window of 2048 samples, while the bottom row is for 1024 samples. The left column is for the 3/4-overlap condition, and the right column is for the half-overlap.

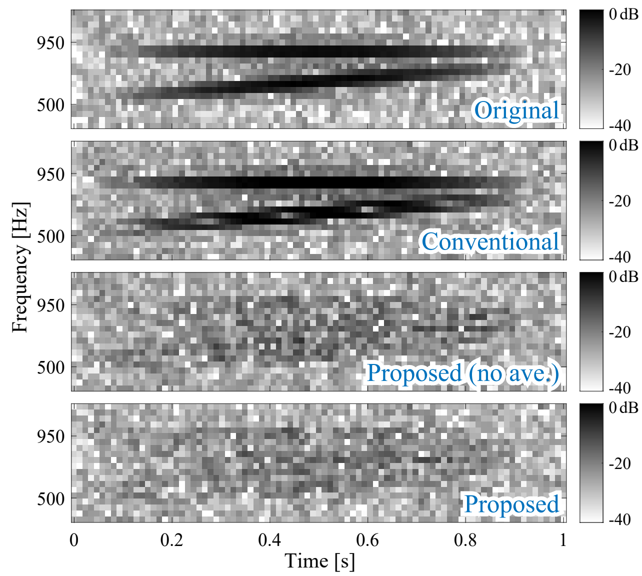


Fig. 3. Magnitude of the phase corrected difference of a noisy signal. SNR was set to 0 dB by adding noise in time domain. The window length and shifting step were 1024 and 512 samples, respectively.

original paper of PCTV (see Fig. 2 of [1]). Since PCTV and iPCTV are summation of the values of all time-frequency bins of the corresponding figure, brighter figure means a better prior which selectively ignore sinusoidal components. That is, such prior can penalizes non-sinusoidal components more effectively. As shown in Fig. 1, the proposed phase correction resulted in the brighter figures than the conventional one, which indicates that the proposed prior was able to correct the frequency mismatch appropriately.

Fig. 2 shows the phase corrected difference for several situations. Every axis and color map are the same as Fig. 1. Since the conventional PCTV only corrects the constant phase as in Eq. (4), it cannot handle these large shifting steps well. However, many applications utilize such combination of the parameters: half-overlap or 3/4-overlap window. A smaller shifting step ends up with high redundancy which may not be preferable in audio applications. On the other hand, the proposed iPCTV can handle those large shifting steps by the additional phase correction calculated from the signal. As shown in the bottom row of Fig. 2, the conventional PCTV also cannot handle the spectral leakage, while the proposed iPCTV can handle it to some extent. Note that iPCTV yielded slightly better results, around the time-varying component, than its variant using Eq. (13) (no averaging). This is because the averaging in Eq. (12) approximates the instantaneous frequency at the midpoint $\delta(m, n-1/2)$ which follows the time-varying component better than $\delta(m,n)$.

The phase corrected difference of a noisy test signal is shown in Fig. 3. The parameters were the same as Fig. 2 (d), and the signal-to-noise ratio (SNR) was set to 0 dB by adding Gaussian random noise in time domain. Since the instantaneous phase utilized in the proposed iPCTV is merely an estimated value, it should contain estimation error caused by the noise. Nevertheless, the proposed iPCTV can effectively ignore the sinusoidal components while it leaves the noise. Therefore, minimizing the proposed prior leads to reduction of non-sinusoidal components as shown in the next section.

3. PHASE-AWARE AUDIO DENOISING BY PCTVS

In order to demonstrate the effectiveness of the proposed prior, the following simple denoising problem is considered in this section: Finding

$$x^\star = \arg\min_{x} \left[\frac{1}{2} \| x - d \|_2^2 + \lambda \| D_t E_{\text{iPC}} \mathscr{F}_{\text{PC}}^{w} x \|_1 \right], \tag{14}$$

where d is the noisy data to be denoised, $\lambda > 0$ is a regularization parameter, and the second term is the proposed iPCTV. The denoising performance is compared with the one using the conventional PCTV:

$$x^\star = \arg\min_{x} \left[\frac{1}{2} \| x - d \|_2^2 + \lambda \| D_t \mathscr{F}_{\text{PC}}^{w} x \|_1 \right], \tag{15}$$

where the only difference is the instantaneous phase correction by E_{iPC}. These problems are strongly convex which ensures the existence of an unique solution. Therefore, many convex optimization algorithms can solve them without pain. In this paper, the primal-dual splitting algorithm is applied.

3.1. Denoising algorithm based on primal-dual splitting

The primal-dual splitting method [11] is an iterative algorithm for finding

$$x^\star \in \arg\min_{x} \Big[f(x) + g(x) + h(\Phi x) \Big], \tag{16}$$

where f is a differential convex function with a β-Lipschitzian gradient, g and h are proper lower-semicontinuous convex functions, and Φ is a bounded linear operator. Its special case can be written as the following procedure [23]: Iterate

$$\left\lfloor \begin{array}{l} \xi^{[k+1]} = \operatorname{prox}_{\sigma_1 g} \big[\xi^{[k]} - \sigma_1 (\nabla f(\xi^{[k]}) + \Phi^* \zeta^{[k]}) \big], \\ \zeta^{[k+1]} = \operatorname{prox}_{\sigma_2 h^*} \big[\zeta^{[k]} + \sigma_2 \Phi (2\xi^{[k+1]} - \xi^{[k]}) \big], \end{array} \right. \tag{17}$$

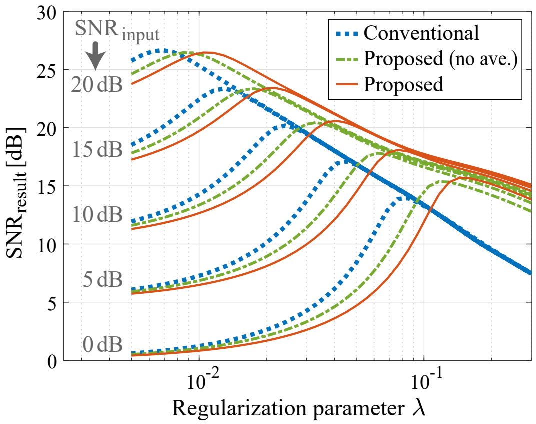


Fig. 4. SNR of the denoised results for each λ. SNR of the noisy speech ($\mathrm{SNR_{input}}$) was adjusted to $\{0, 5, 10, 15, 20\}$ by adding the Gaussian random noise in time domain.

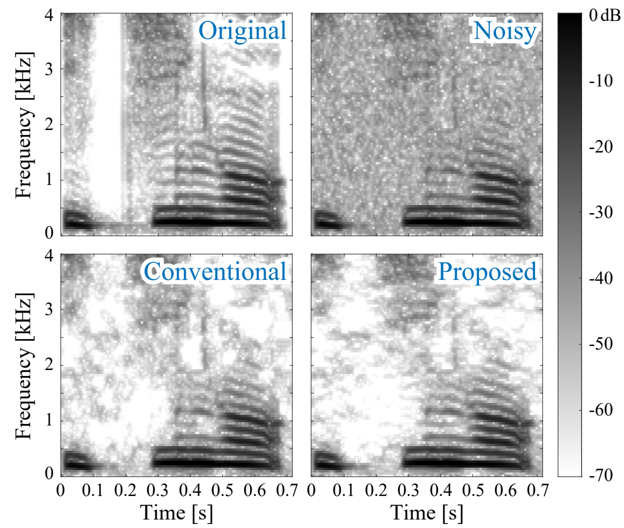


Fig. 5. An example of the denoised spectrograms for $\mathrm{SNR_{input}} = 20$ dB. Based on Fig. 4, the regularization parameters λ were chosen so that their best $\mathrm{SNR_{result}}$ were obtained for each prior. The PESQ scores corresponding to the above spectrograms were 2.346 (noisy), 3.105 (conventional), and 3.206 (proposed).

where k is iteration index, Φ^* is the adjoint of Φ, $\sigma_1, \sigma_2 > 0$,

$$\mathrm{prox}_{\lambda f}[z] = \arg\min_x \Big[\, f(x) + \frac{1}{2\lambda}\|z - x\|_2^2 \,\Big] \tag{18}$$

is the proximity operator, and h^* is the convex conjugate of h whose proximity operator can be calculated as

$$\mathrm{prox}_{\sigma h^*}[z] = z - \sigma\, \mathrm{prox}_{h/\sigma}[z/\sigma] \tag{19}$$

for $\sigma > 0$. By letting $f(x) = \frac{1}{2}\|x - d\|_2^2$, $g(x) = 0$, $h(x) = \lambda\|x\|_1$, and $\Phi = D_t E_{\mathrm{iPC}} \mathscr{F}^w_{\mathrm{PC}}$, Eq. (17) yields the following algorithm which solves Eq. (14):

$$\left\lfloor \begin{array}{l} x^{[k+1]} = x^{[k]} - \sigma_1(x^{[k]} - d + \mathscr{F}^{w*}_{\mathrm{PC}} E^*_{\mathrm{iPC}} D^*_t z^{[k]}), \\ z^{[k+1]} = \widetilde{\mathcal{T}}_\lambda\big[\, z^{[k]} + \sigma_2 D_t E_{\mathrm{iPC}} \mathscr{F}^w_{\mathrm{PC}}(2x^{[k+1]} - x^{[k]})\big], \end{array} \right. \tag{20}$$

where $\mathscr{F}^{w*}_{\mathrm{PC}}$ is the inverse STFT using the window w, $(E^*_{\mathrm{iPC}} z)(m,n) = z(m,n)e^{2\pi i b a \tilde{\delta}(m,n)/L}$, D^*_t is the time-directional difference in the opposite direction of D_t [3], and

$$\big(\widetilde{\mathcal{T}}_\lambda[z]\big)_n = \min\{1, \lambda/|z_n|\} z_n. \tag{21}$$

The convergence of this algorithm to the unique solution of Eq. (14) is guaranteed when σ_1 and σ_2 satisfies [23]

$$\frac{1}{\sigma_1} - \sigma_2 \,\|D_t E_{\mathrm{iPC}} \mathscr{F}^w_{\mathrm{PC}}\|^2_{\mathrm{op}} \geq \frac{1}{2}, \tag{22}$$

where $\|\cdot\|_{\mathrm{op}}$ denotes the operator norm. Note that the algorithm in Eq. (20) can also solve Eq. (15) by just omitting E_{iPC} and E^*_{iPC}.

3.2. Result of speech denoising

A female speech signal, whose sampling frequency was 44100 Hz, was utilized to test the performance of the priors. STFT was calculated by the Hann window of 1024 samples with 256 sample shifting.

Fig. 4 shows denoising performance in terms of SNR measured in time domain. $\mathrm{SNR_{input}}$ denotes SNR of the noisy speech to be denoised, and $\mathrm{SNR_{result}}$ is that of the denoised signals. The regularization parameter λ was chosen as in the horizontal axis to illustrate its effect on the performance. As in the figure, the proposed method improved the denoising performance especially for the low SNR condition. Moreover, the proposed iPCTV is relatively insensitive to the value of λ comparing to the conventional PCTV because the width of the curve of iPCTV is wider than that of PCTV. Therefore, parameter tuning of iPCTV is possibly easier than that of the conventional PCTV.

Fig. 5 shows an example of denoised spectrograms for $\mathrm{SNR_{input}} = 20$ dB, which is the condition where the difference between the priors cannot be seen in terms of $\mathrm{SNR_{result}}$. The regularization parameters λ were chosen according to Fig. 4 so that both PCTV and iPCTV achieved their best $\mathrm{SNR_{result}}$. In the figures, the proposed prior seems to obtain a slightly better result than the conventional one as the harmonic component around 0.4 s is smoother. In addition, the proposed iPCTV has less remaining noise around 0.1–0.2 s. Such differences were not apparent in terms of SNR, but they can be noted as a score of PESQ (Perceptual Evaluation of Speech Quality [24]) which is the standard speech quality assessment method. The score of PESQ was 2.346 for the noisy signal, and 3.105 and 3.206 for the conventional and proposed proiors, respectively. The improvement of PESQ for the conventional PCTV was 0.759 while that for the proposed iPCTV was 0.860. These results suggest that the proposed prior is a successful improvement of the simple prior, PCTV, without increasing the computational complexity much.

4. CONCLUSION

In this paper, a simple prior for audio signals is proposed. It is based on the conventional PCTV which is the time-directional total variation combined with the phase correction. The proposed prior, iPCTV, improves PCTV by the instantaneous phase correction so that the phase spectrogram is corrected based on the signal's phase. The denoising experiment showed the potentialities of the proposed method. Since the proposed iPCTV is simple and effective, it can be a building block of a processing method for many audio applications. Investigating such signal processing methods (might be based upon existing ones [25–35]) is remained as the future work.

5. REFERENCES

[1] İ. Bayram and M. E. Kamasak, "A simple prior for audio signals," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 6, pp. 1190–1200, Jun. 2013.

[2] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, no. 1-4, pp. 259–268, 1992.

[3] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, no. 1, pp. 89–97, 2004.

[4] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM J. Imag. Sci.*, vol. 3, no. 3, pp. 492–526, 2010.

[5] S. Ono and I. Yamada, "Decorrelated vectorial total variation," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 4090–4097.

[6] T. Gerkmann, M. Krawczyk-Becker, and J. Le 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.

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

[8] P. L. Combettes and J. C. Pesquet, *Proximal Splitting Methods in Signal Processing*, pp. 185–212, Springer New York, New York, NY, 2011.

[9] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[10] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, 2014.

[11] N. Komodakis and J. C. Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," *IEEE Signal Process. Mag.*, vol. 32, no. 6, pp. 31–54, Nov. 2015.

[12] S. Ono and I. Yamada, "Hierarchical convex optimization with primal-dual splitting," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 373–388, Jan. 2015.

[13] H. G. Feichtinger and T. Strohmer, Eds., *Gabor Analysis and Algorithms: Theory and Applications*, Birkhäuser Boston, Boston, MA, 1998.

[14] K. Gröchenig, *Foundations of Time-Frequency Analysis*, Birkhäuser Boston, Boston, MA, 2001.

[15] P. L. Søndergaard, "Gabor frames by sampling and periodization," *Adv. Comput. Math.*, vol. 27, no. 4, pp. 355–373, 2007.

[16] K. Kodera, C. De Villedary, and R. Gendrin, "A new method for the numerical analysis of non-stationary signals," *Phys. Earth Planet. Interiors*, vol. 12, no. 2, pp. 142–150, 1976.

[17] K. Kodera, R. Gendrin, and C. De Villedary, "Analysis of time-varying signals with small BT values," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 26, no. 1, pp. 64–76, Feb. 1978.

[18] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1068–1089, May. 1995.

[19] S. A. Fulop and K. Fitz, "Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications," *J. Acoust. Soc. Am.*, vol. 119, no. 1, pp. 360–371, 2006.

[20] 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, Nov. 2013.

[21] N. Holighaus, Z. Průša, and P. L. Søndergaard, "Reassignment and synchrosqueezing for general time-frequency filter banks, subsampling and processing," *Signal Process.*, vol. 125, pp. 1–8, 2016.

[22] S. Fenet, R. Badeau, and G. Richard, "Reassigned time-frequency representations of discrete time signals and application to the Constant-Q Transform," *Signal Process.*, vol. 132, pp. 170–176, 2017.

[23] L. Condat, "A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms," *J. Opt. Theory Appl.*, vol. 158, no. 2, pp. 460–479, 2013.

[24] ITU-T Rec. P. 862, *Perceptual evaluation of speech quality (PESQ)*, 2000.

[25] K. Yatabe and Y. Oikawa, "Optically visualized sound field reconstruction based on sparse selection of point sound sources," in *IEEE ICASSP*, Apr. 2015, pp. 504–508.

[26] T. Tachikawa, K. Yatabe, and Y. Oikawa, "Coherence-adjusted monopole dictionary and convex clustering for 3D localization of mixed near-field and far-field sources," in *IEEE ICASSP*, Mar. 2017, pp. 3191–3195.

[27] K. Yatabe and Y. Oikawa, "Convex optimization-based windowed Fourier filtering with multiple windows for wrapped-phase denoising," *Appl. Opt.*, vol. 55, no. 17, pp. 4632–4641, Jun. 2016.

[28] Y. Tamura, K. Yatabe, and Y. Oikawa, "Least-squares estimation of sound source directivity using convex selector of a better solution," *Acoust. Sci. Tech.*, vol. 38, no. 3, pp. 128–136, 2017.

[29] K. Yatabe and Y. Oikawa, "PDE-based interpolation method for optically visualized sound field," in *IEEE ICASSP*, May 2014, pp. 4738–4742.

[30] Y. Koyano, K. Yatabe, and Y. Oikawa, "Infinite-dimensional SVD for revealing microphone array's characteristics," *Appl. Acoust.*, vol. 129, pp. 116–125, 2018.

[31] K. Yatabe and Y. Oikawa, "Optically visualized sound field reconstruction using Kirchhoff–Helmholtz equation," *Acoust. Sci. Tech.*, vol. 36, no. 4, pp. 351–354, 2015.

[32] N. Chitanont, K. Yatabe, K. Ishikawa, and Y. Oikawa, "Spatio-temporal filter bank for visualizing audible sound field by Schlieren method," *Appl. Acoust.*, vol. 115, pp. 109–120, 2017.

[33] K. Yatabe, K. Ishikawa, and Y. Oikawa, "Acousto-optic back-projection: Physical-model-based sound field reconstruction from optical projections," *J. Sound Vib.*, vol. 394, pp. 171–184, 2017.

[34] Y. Masuyama, T. Kusano, K. Yatabe, and Y. Oikawa, "Modal decomposition of musical instrument sound via alternating direction method of multipliers," in *IEEE ICASSP*, Apr. 2018.

[35] K. Yatabe and D. Kitamura, "Determined blind source separation via proximal splitting algorithm," in *IEEE ICASSP*, Apr. 2018.