Phase Corrected Total Variation for Audio Signals

Kohei Yatabe and **Yasuhiro Oikawa** (Waseda University)

Introduction

- In optimization-based signal processing, the **prior term** $\mathcal{P}$ models the desired signal, while the **data term** $\mathcal{D}$ models the observation process.

$$\text{Find}\ \ x^\star \in \arg\min_{x} \big[\, \mathcal{D}(x,d) + \lambda\,\mathcal{P}(x) \,\big]$$

- **Design of prior terms is crucial to the quality of processing**. Some requirements of a good prior term are (1) optimization friendliness, (2) computational easiness, and (3) model preciseness.
- **Total Variation (TV)**, which is a popular prior in image processing, has been **applied to complex spectrograms with phase correction** as a computationally and optimization friendly **phase-aware prior** [1].
- The simple prior proposed in Ref. [1], what we call Phase Corrected Total Variation (PCTV), has some **model mismatch in terms of phase** that might restrict its effectiveness to some extent.
- In this paper, an **improved version of PCTV** based on **instantaneous phase correction (iPCTV)** is proposed to reduce the model mismatch.

Phase Corrected Total Variation (PCTV) [1]

- The short-time Fourier transform (STFT) of a sinusoidal signal has the **neighborhood relation** when its frequency **coincides with some bin**.

$$(\mathscr{F}^w s)[t+1,f]\, e^{-2\pi \mathrm{i} b f a/L} = (\mathscr{F}^w s)[t,f]$$

$$(\mathscr{F}^w x)[t,f] = \sum_{l=0}^{L-1} x[l+at]\, \overline{w[l]\, e^{2\pi \mathrm{i} b f l/L}}$$

- **PCTV** was defined as the **time-directional Total Variation [1]** where the **Phase Correction (PC)** was utilized for canceling the phase factor.

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

$$(D_t x)[t,f] = x[t,f] - x[t-1,f], \qquad (E_{\mathrm{PC}} x)[t,f] = x[t,f]\, e^{-2\pi \mathrm{i} b f a t/L}$$

$$(\mathscr{F}^w_{\mathrm{PC}} x)[t,f] = \sum_{l=0}^{L-1} x[l]\, \overline{w[l-at]\, e^{2\pi \mathrm{i} b f l/L}}$$

- The phase correction is **valid only when** frequencies of components contained in the signal **coincide with the frequencies of bins** of STFT.

Proposed Prior (iPCTV) Improved by Instantaneous Phase Correction (iPC)

- The neighborhood relation of a sinusoidal signal has an **additional phase factor** when the frequency does not coincide with the bin.

$$(\mathscr{F}^w \tilde{s})[t+1,f]\, e^{-2\pi \mathrm{i} b f a/L}\, e^{-2\pi \mathrm{i} b \delta a/L} = (\mathscr{F}^w \tilde{s})[t,f]$$

- Phase correction can be performed correctly for any sinusoidal signals **if the frequency mismatch** δ **is known** in advance. This mismatch factor can be regarded as an approximated **instantaneous frequency**.

$$\delta[t,f] = -\,\mathrm{Im}\big\{\, (\mathscr{F}^{w'}_{\mathrm{PC}} x)[t,f] \,/\, (\mathscr{F}^{w}_{\mathrm{PC}} x)[t,f] \,\big\} \,/\, b$$

- This instantaneous frequency **can be calculated from observed data** that opens the possibility of correcting phase without the mismatch.
- We propose an improved version of PCTV, what we call **iPCTV**, based on the **instantaneous phase correction (iPC)** of the spectrogram.

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

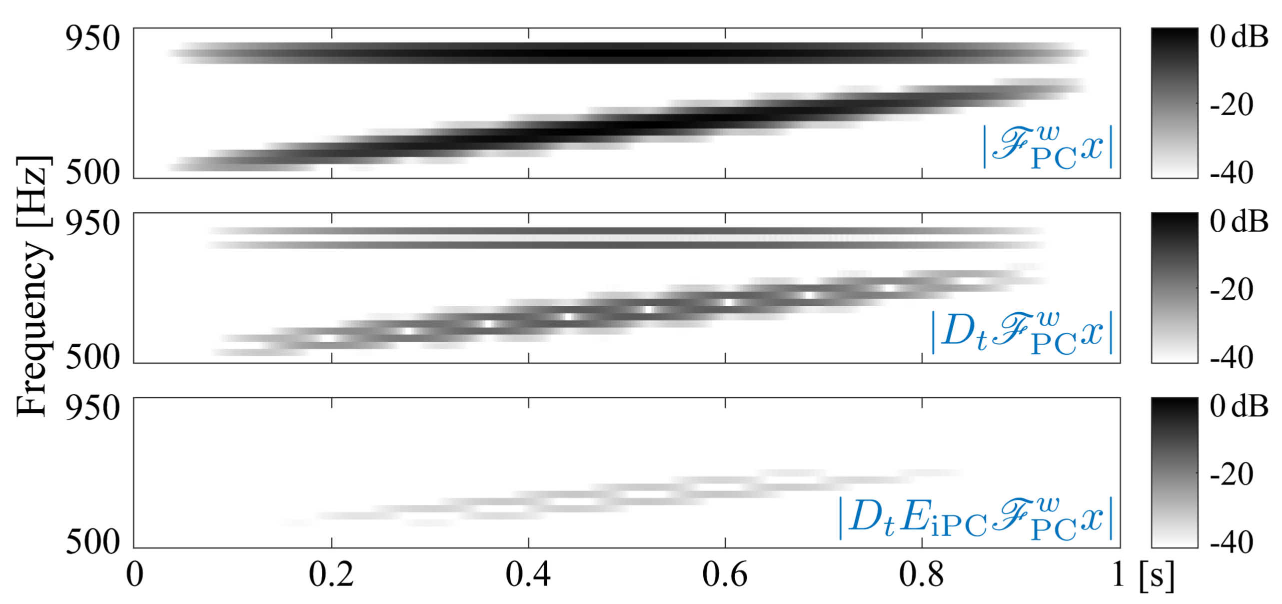
$$(E_{\mathrm{iPC}} x)[t,f] = x[t,f]\, e^{-2\pi \mathrm{i} b a \tilde{\delta}[t,f]/L}$$

- Instantaneous phase $\tilde{\delta}$ is calculated from the instantaneous frequency only once so that **convexity of PCTV is conserved**.

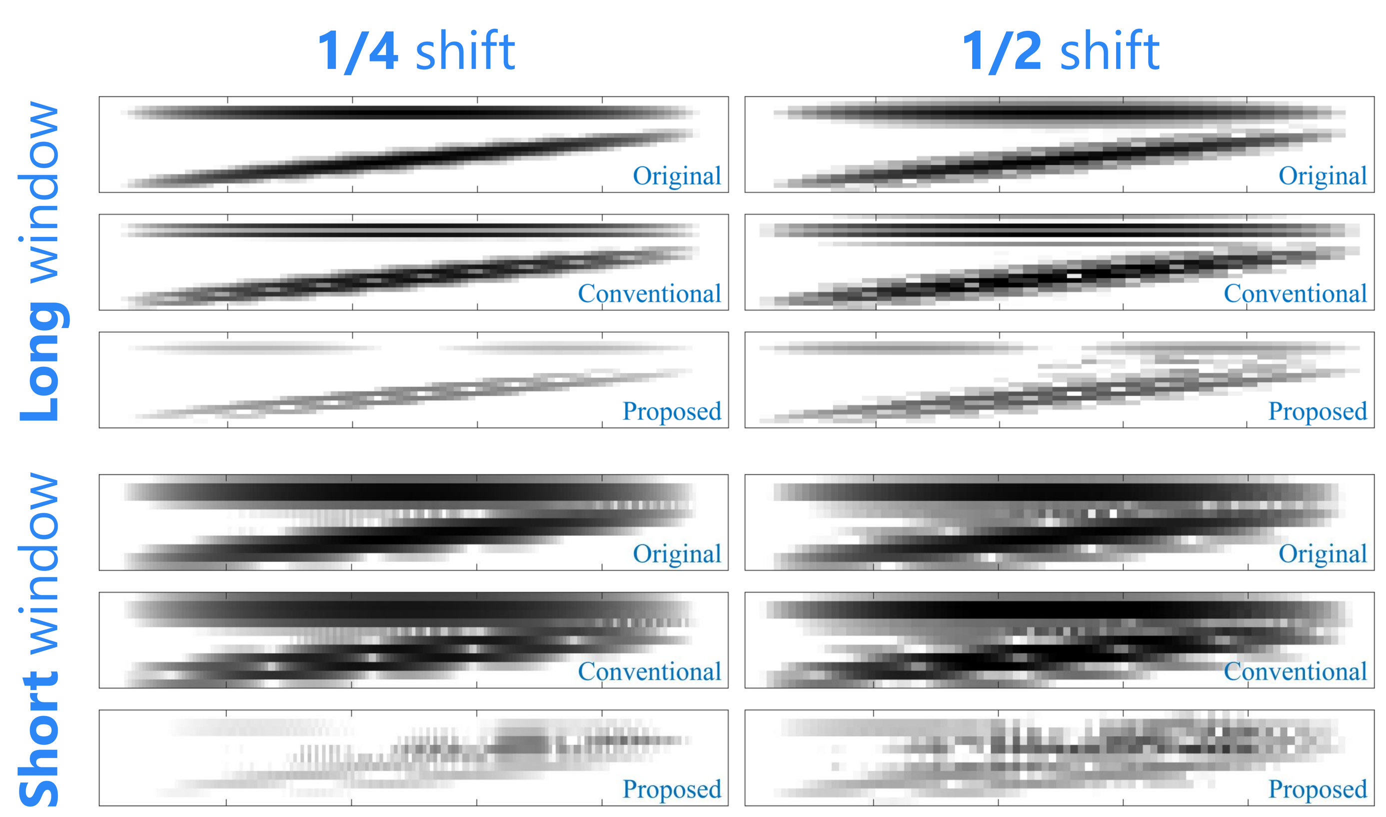
$$\tilde{\delta}[t,f] = \begin{cases} \sum_{l=0}^{t-1} \big(\, \delta[l+1,f] + \delta[l,f] \,\big) \,/\, 2 & (t \geq 1) \\ 0 & (t = 0) \end{cases}$$

Comparison of the Conventional PCTV and the Proposed iPCTV

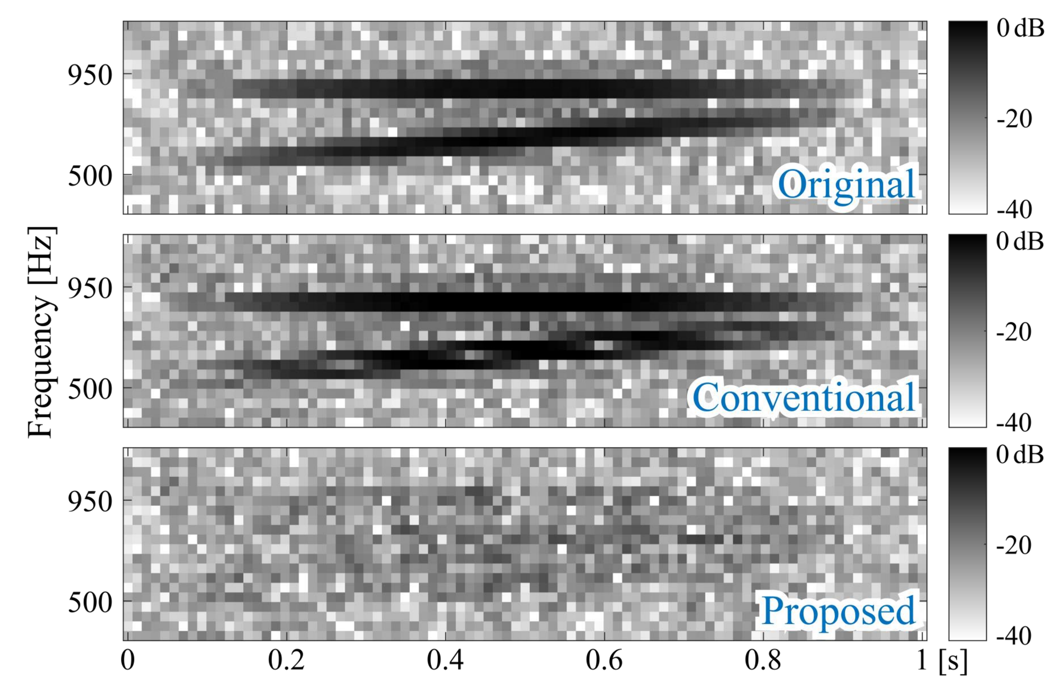
- The proposed **iPCTV can attenuate sinusoidal components more effectively** than the conventional PCTV. (long window, 1/16 shift)



- The proposed **iPCTV can handle wider window sifting width**, while the performance of the conventional PCTV deteriorates for wider shift.

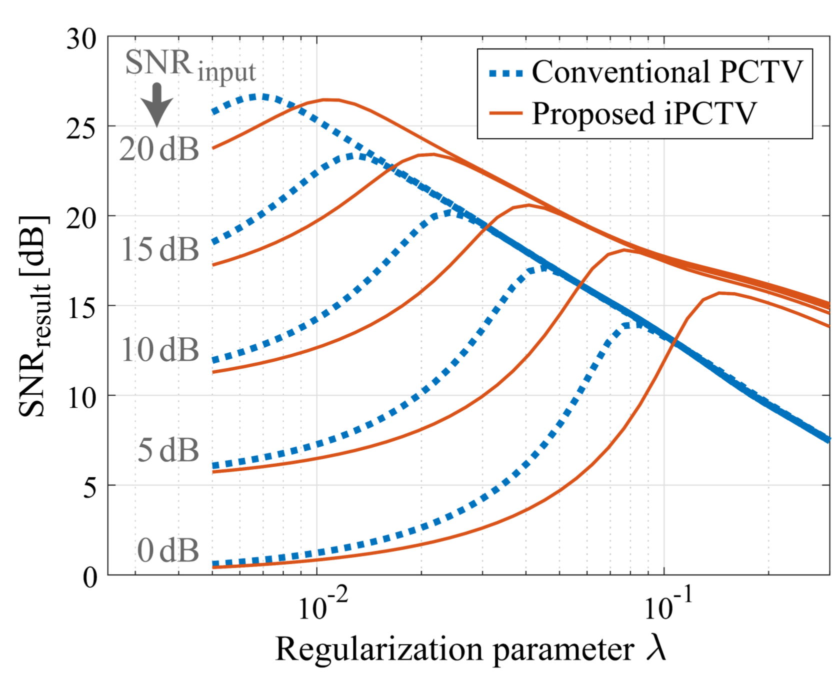


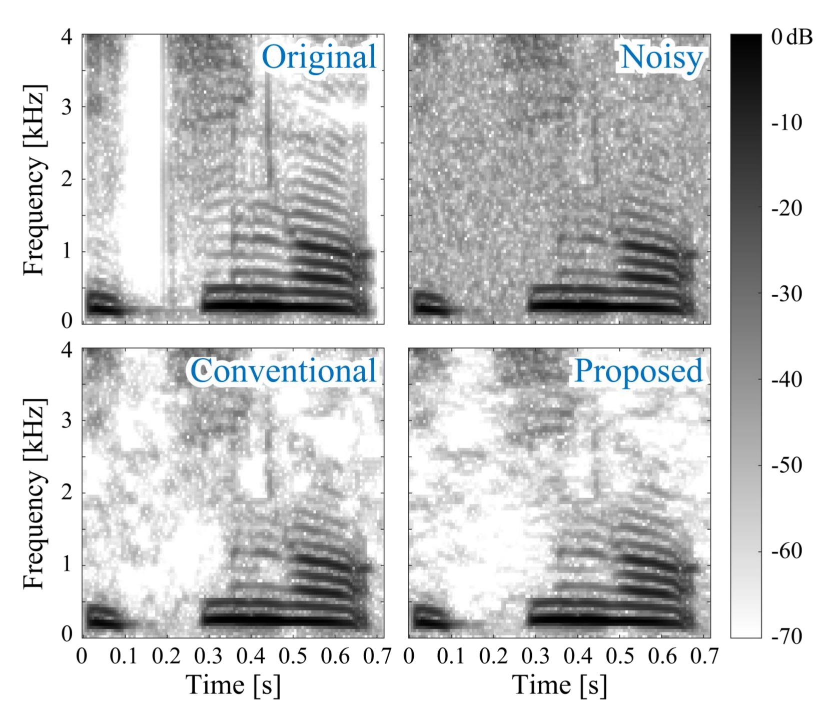
- The proposed instantaneous phase correction **works correctly even when its phase factor is calculated from noisy observation**.



- Simple speech denoising performance of iPCTV was better than PCTV.

$$\text{Find}\ \ x^\star \in \arg\min_{x} \big[\, \|x-d\,\|_2^2 \,/\, 2 + \lambda\, \mathrm{TV}_{\mathrm{iPC}}(x) \,\big]$$





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