Phase Corrected Total Variation for Audio Signals

Kohei Yatabe and Yasuhiro Oikawa (Waseda University)

Introduction

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

\[
    \text{Find } x^* \in \arg \min_x [D(x, d) + \lambda P(x)]
\]

- 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.

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.

\[
    (\mathcal{F}^{w}\hat{s})[t + 1, f] e^{-2\pi ibf/t/L} e^{-2\pi ibf_0/L} = (\mathcal{F}^{w}\hat{s})[t, f]
\]

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

\[
    \delta[t, f] = -\text{Im}\{ (\mathcal{F}^{w}_{\text{iPC}}x)[t, f] / (\mathcal{F}^{w}_{\text{iPC}}x)[t, f] \} / b
\]

- This instantaneous frequency can be calculated from observed data that opens the possibility of correcting phase without the mismatch.

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.

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.

\[
    (\mathcal{F}^{w}s)[t + 1, f] e^{-2\pi ibf/t/L} = (\mathcal{F}^{w}s)[t, f]
\]

\[
    (\mathcal{F}^{w}x)[t, f] = \sum_{l=0}^{L-1} x[t + a[l] w[t] e^{2\pi ibf/L}
\]

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

\[
    \text{TV}_{\text{PC}}(x) = \| D_i \mathcal{F}^{w}_{\text{PC}}x \|_1 = \| D_i E_{\text{PC}} \mathcal{F}^{w}x \|_1
\]

\[
    (E_{\text{PC}}x)[t, f] = x[t, f] e^{-2\pi ibf\delta[t, f]/L}
\]

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

We propose an improved version of PCTV, what we call iPCTV, based on the instantaneous phase correction (iPC) of the spectogram.

\[
    \text{TV}_{\text{iPC}}(x) = \| D_i E_{\text{iPC}} \mathcal{F}^{w}_{\text{iPC}}x \|_1 = \| D_i E_{\text{iPC}} E_{\text{PC}} \mathcal{F}^{w}x \|_1
\]

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

\[
    \tilde{\delta}[t, f] = \left\{ \begin{array}{ll}
    \sum_{l=0}^{t-1} (\delta[l + 1, f] + \delta[l, f]) / 2 & (t \geq 1) \\
    0 & (t = 0)
    \end{array} \right.
\]

- 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.

\[
    \text{Find } x^* \in \arg \min_x [\| x - d \|_2^2 / 2 + \lambda \text{TV}_{\text{iPC}}(x)]
\]