WASEDA University
早稲田大学

Phase-aware harmonic/percussive source separation via convex optimization

Yoshiki Masuyama*, Kohei Yatabe, Yasuhiro Oikawa (Waseda University, Japan) *Email: mas-03151102@akane.waseda.jp

Introduction

- **Harmonic/percussive source separation (HPSS)**, is a useful pre-processing tool for many audio applications.
- **Conventional**: The distinctive source-specific structures of amplitude spectrograms has been used.
- **Problem**: Phase reconstruction is required because of their amplitude-only treatment.
- **Proposal**: **A optimization-based HPSS method** simultaneously **treats the amplitude and phase**.
- **Results**: The numerical experiment validated the effectiveness of the proposed method in terms of SDR.

HPSS based on anisotropic smoothness

- HPSS based on anisotropic smoothness assumes the power spectrograms of harmonic and percussive components have the following relation:

$$H_{\omega,\tau} \approx H_{\omega,\tau\pm 1}, \quad P_{\omega,\tau} \approx P_{\omega\pm 1,\tau}$$

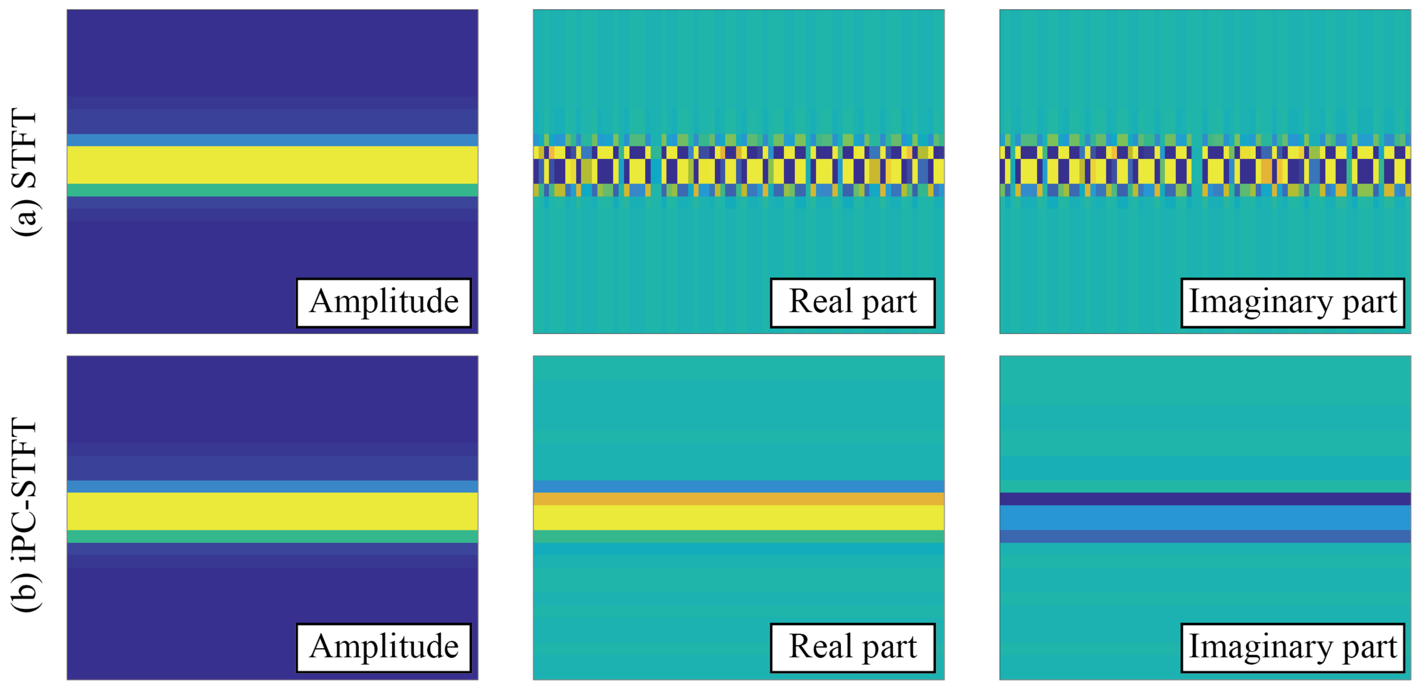
- Based on the above relation, a HPSS method is formulated as the following optimization problem.

$$\min_{\mathbf{H},\mathbf{P}} \ \frac{1}{2\sigma_{\mathrm{h}}^2}\|D_\tau(\mathbf{H})\|_{\mathrm{Fro}}^2 + \frac{1}{2\sigma_{\mathrm{p}}^2}\|D_\omega(\mathbf{P})\|_{\mathrm{Fro}}^2$$
$$\text{s.t.} \quad H_{\omega,\tau} + P_{\omega,\tau} = |X_{\omega,\tau}|^{2\gamma}, \quad H_{\omega,\tau} \geq 0, \quad P_{\omega,\tau} \geq 0$$

- This approach considers **only amplitude spectrograms**, and thus **the phase information is ignored**.

Property of complex-valued spectrograms

- Although the amplitude spectrogram of **a sinusoid** is constant in the time direction, its **complex-valued spectrogram periodically fluctuates.**



Proposed phase-aware HPSS method

Instantaneous phase corrected total variation

- The complex-valued spectrogram of a sinusoid has the following relation.

$$\mathscr{F}(\mathbf{x})_{\omega,\tau} = \mathscr{F}(\mathbf{x})_{\omega,\tau-1}\, \mathrm{e}^{2\pi j f a/L}$$

- The complex-valued spectrogram becomes constant in each sub-band if its phase evolution is eliminated.
- We use **the instantaneous phase corrected STFT (iPC-STFT)** proposed in our previous study.

$$\mathscr{F}_{\mathrm{iPC}}(\mathbf{x})_{\omega,\tau} = \prod_{\eta=0}^{\tau-1} \mathrm{e}^{-2\pi j v_{\omega,\eta} a/L}\, \mathscr{F}(\mathbf{x})_{\omega,\tau}$$

- **The iPC-STFT spectrogram of a sinusoid is smooth in each sub-band** when the instantaneous frequency is accurately estimated.
- The instantaneous frequency can be estimated from the observed signal.

$$v_{\omega,\tau} = b\omega - \mathrm{Im}[\tilde{\mathscr{F}}(\mathbf{x})_{\omega,\tau} / \mathscr{F}(\mathbf{x})_{\omega,\tau}]$$

Proposed optimization-based HPSS method

- Using iPC-STFT, we propose **a phase-aware HPSS method** through the following **convex optimization**.

$$\min_{\mathbf{x}_{\mathrm{h}},\mathbf{x}_{\mathrm{p}}} \ \frac{1}{2}\|\mathbf{W} \odot D_\tau(\mathbf{X}_{\mathrm{h}})\|_{\mathrm{Fro}}^2 + \lambda\|\mathbf{X}_{\mathrm{p}}\|_{2,1}$$
$$\text{s.t.} \quad \mathbf{x} = \mathbf{x}_{\mathrm{h}} + \mathbf{x}_{\mathrm{p}}, \quad \mathbf{X}_{\mathrm{h}} = \mathscr{F}_{\mathrm{iPC}}(\mathbf{x}_{\mathrm{h}}), \quad \mathbf{X}_{\mathrm{p}} = \mathscr{F}(\mathbf{x}_{\mathrm{p}})$$

$\mathbf{x}$: the time domain observed signal	λ : a regularization parameter
$\mathbf{x}_{\mathrm{h}}$: the time domain signal of harmonic components	$\mathbf{W}$: a data-dependent weight
$\mathbf{x}_{\mathrm{p}}$: the time domain signal of percussive components	

- Main points of the proposed method:
  - **The phase-aware smoothness** in the time direction is assumed for **harmonic components**.
  - **The time-frame-wise sparsity** is considered for **percussive components** instead of the frequency-directional smoothness.
  - It treats **variables in the time domain**, and the perfect reconstruction constraint is introduced.
  - It is formulated as **a convex optimization** problem.

Proposed algorithm

Primal-dual splitting algorithm

- Primal-dual splitting (PDS) algorithm is one of the proximal splitting algorithms which can solve the following convex optimization problem.

$$\min_{\mathbf{x}} \ \Theta(\mathbf{x}) + \Upsilon_1(\mathscr{L}_1(\mathbf{x})) + \Upsilon_2(\mathscr{L}_2(\mathbf{x}))$$

- To apply the PDS algorithm to the proposed HPSS problem, we reformulate it as follows.

$$\min_{\mathbf{x}_{\mathrm{h}},\mathbf{x}_{\mathrm{p}}} \ \iota_{\mathbf{x}}(\mathbf{x}_{\mathrm{h}},\mathbf{x}_{\mathrm{p}}) + \frac{1}{2}\|\mathscr{L}_{\mathrm{h}}(\mathbf{x}_{\mathrm{h}})\|_{\mathrm{Fro}}^2 + \lambda\|\mathscr{F}(\mathbf{x}_{\mathrm{p}})\|_{2,1}$$

Algorithm 1 Proposed HPSS algorithm

Input: $\mathbf{x}, \mathbf{x}_{\mathrm{h}}^{[0]}, \mathbf{x}_{\mathrm{p}}^{[0]}, \mathbf{Y}_{\mathrm{h}}^{[0]}, \mathbf{Y}_{\mathrm{p}}^{[0]}, \lambda, \mu_1, \mu_2, \alpha$
Output: $\mathbf{x}_{\mathrm{h}}^{[n+1]}, \mathbf{x}_{\mathrm{p}}^{[n+1]}$
for $n = 1, 2, \ldots$ **do**
$(\tilde{\mathbf{x}}_{\mathrm{h}}, \tilde{\mathbf{x}}_{\mathrm{p}}) = P_{\mathbf{x}}\big(\mathbf{x}_{\mathrm{h}}^{[n]} - \mu_1\mathscr{L}_{\mathrm{h}}^*(\mathbf{Y}_{\mathrm{h}}^{[n]}),\ \mathbf{x}_{\mathrm{p}}^{[n]} - \mu_1\mathscr{F}^*(\mathbf{Y}_{\mathrm{p}}^{[n]})\big)$
$\mathbf{z}_{\mathrm{h}} = \mathbf{y}_{\mathrm{h}}^{[n]} + \mathscr{L}_{\mathrm{h}}(2\tilde{\mathbf{x}}_{\mathrm{h}} - \mathbf{x}_{\mathrm{h}}^{[n]})$
$\mathbf{z}_{\mathrm{p}} = \mathbf{y}_{\mathrm{p}}^{[n]} + \mathscr{F}(2\tilde{\mathbf{x}}_{\mathrm{p}} - \mathbf{x}_{\mathrm{p}}^{[n]})$
$\tilde{\mathbf{y}}_{\mathrm{h}} = \mathbf{z}_{\mathrm{h}} - \mu_2\,\mathrm{prox}_{(1/\mu_2)\|\cdot\|_{\mathrm{Fro}}^2}(\mathbf{z}_{\mathrm{h}}/\mu_2)$
$\tilde{\mathbf{y}}_{\mathrm{p}} = \mathbf{z}_{\mathrm{p}} - \lambda\mu_2\,\mathrm{prox}_{(1/\lambda\mu_2)\|\cdot\|_{2,1}}(\mathbf{z}_{\mathrm{p}}/\lambda\mu_2)$
$(\mathbf{x}_{\mathrm{h,p}}^{[n+1]}, \mathbf{y}_{\mathrm{h,p}}^{[n+1]}) = \alpha(\tilde{\mathbf{x}}_{\mathrm{h,p}}, \tilde{\mathbf{y}}_{\mathrm{h,p}}) + (1-\alpha)(\mathbf{x}_{\mathrm{h,p}}^{[n]}, \mathbf{y}_{\mathrm{h,p}}^{[n]})$
end for

Numerical experiment

- The proposed method was compared with the anisotropic smoothness based method (Ono's), median-filtering (MF), kernel additive model (KAM), and phase-aware time-frequency masking (PM).
- **Prop-mix** (using instantaneous frequency estimated from the noisy signal) **outperformed the conventional methods in terms of SDR**, and Prop-ora achieved the highest SDR.

Experimental condition	
Sampling rate	44100 Hz
Window	Hann
Window length	93 ms
Shift length	23 ms
λ	0.5

		Ono's	MF	KAM	PM	Prop-mix	Prop-ora
Har.	SDR	5.8	8.6	4.9	−8.6	**9.3**	10.3
	SIR	11.2	15.1	**23.1**	6.1	12.3	13.8
	SAR	7.6	10.2	5.1	−7.5	**15.4**	15.4
Per.	SDR	−8.1	−4.2	−4.7	−12.1	**−3.8**	−2.7
	SIR	−2.8	−1.3	−2.3	−3.2	**1.7**	2.8
	SAR	−1.9	3.5	**4.2**	−6.7	1.6	2.6
Ave.	SDR	−1.1	2.0	0.1	−10.4	**2.8**	3.8
	SIR	4.2	6.9	**10.4**	1.5	7.0	8.3
	SAR	2.8	6.9	4.6	−7.1	**8.5**	9.0

Conclusion

- We proposed a phase-aware HPSS method through convex optimization which treats both amplitude and phase simultaneously.
- The experimental results indicated the accurate estimation of the instantaneous frequency can improve the performance of the proposed method, which is included in our future works.