MUSICAL SIGNAL SEPARATION BASED ON BAYESIAN SPECTRAL AMPLITUDE ESTIMATOR WITH AUTOMATIC TARGET PRIOR ADAPTATION

Yuki Murota, Daichi Kitamura, Shunsuke Nakai, Hiroshi Saruwatari, Satoshi Nakamura (Nara Institute of Science and Technology, Nara, Japan)
Kazunobu Kondo, Yu Takahashi (Yamaha Corporation, Shizuoka, Japan)

1. Introduction

- Recently, music signal separation technologies have received much attention.

Applications

- Automatic music transcription
- Sound augmented reality (AR)
- 3D audio system, etc.

■ **Previous research**

- **Generalized minimum mean-square error short-time spectral amplitude (MMSE-STSA) estimator[1], [2]**

The amplitude spectrum of the target signal is enhanced on the basis of the MMSE criterion.

Optimal Bayesian estimators based on the a priori target signal statistical model.

Generalized MMSE-STSA can enhance target signal in time-frequency domain.

$$Y_*(f,\tau)=S_*(f,\tau)+N_*(f,\tau) \quad *=\{\mathrm{R},\mathrm{I}\}$$

$Y_*(f,\tau)$: Observed signal $S_*(f,\tau)$: Target signal $N_*(f,\tau)$: Interference signal
f : Frequency bin τ : Frame index $*=\{\mathrm{R},\mathrm{I}\}$: Real and imaginary parts of the signal

It is difficult to deal with nonstationary interference signals.
Priori statistical model of target signal cannot be determined automatically.

- **Supervised Nonnegative matrix factorization (SNMF) [3], [4]**

Sparse representation and decomposition algorithm.
Use some sample sound of the target instrumental signal in a priori training in NMF.
NMF attempts to separate instrumental sources using spectral characteristics **[5]**
SNMF can deal with nonstationary signals.
The mixture model of NMF approximately assumes the additivity of amplitude spectrums.f

$$|Y(f,\tau)|\simeq|S(f,\tau)|+|N(f,\tau)|$$

Purpose of our research

To cope with the problems of Generalized MMSE-STSA estimator and supervised NMF, we propose a signal separation technique which is based on the right mixture model and can deal with nonstationary interference signals.

2. Generalized MMSE-STSA estimator

- In the generalized MMSE-STSA estimator, the a priori statistical model of the target signal amplitude spectrum is set to chi distribution.

Chi- distribution

$$p(x)=\frac{2}{\Gamma(\rho)}\left(\frac{\rho}{\mathrm{E}[x^2]}\right)^{\rho}x^{2\rho-1}\exp\left(\frac{\rho}{\mathrm{E}\ [x^2]}x^2\right)$$

$\Gamma(\cdot)$: Gamma function ρ : Shape parameter
$p(x)$: p.d.f. of signal x in the amplitude domain

➢ ρ = 1 gives a Rayleigh distribution that corresponds to a Gaussian distribution in the time domain.
➢ A smaller value of ρ corresponds to a supper-Gaussian distribution signal.

- The processed signal $\tilde{S}_*(f,\tau)$ via the generalized MMSE-STSA estimator is given as follows.

Target signal estimation by generalized MMSE-STSA estimator

$$\tilde{S}_*(f,\tau)=G(f,\tau)Y_*(f,\tau)$$

$$G(f,\tau)=\frac{\sqrt{\nu(f,\tau)}}{\gamma(f,\tau)}\cdot\frac{\Gamma(\rho+0.5)}{\Gamma(\rho)}\cdot\frac{\Phi(0.5-\rho,1,-\nu(f,\tau))}{\Phi(1-\rho,1,-\nu(f,\tau))}$$

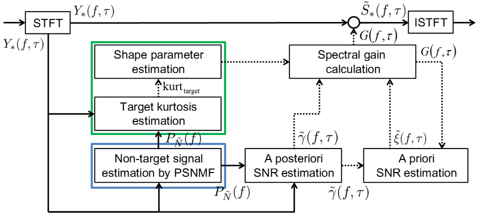
$\tilde{S}_*(f,\tau)$: Estimated target signal $G(f,\tau)$: Gain function
$P_{\tilde{N}}(f)$: Interference signal power spectra α : Forgetting factor
$\Phi(a,b;k)$: Confluent hypergeometric function $\nu(f,\tau)=\tilde{\gamma}(f,\tau)\tilde{\xi}(f,\tau)\left(1+\tilde{\xi}(f,\tau)\right)^{-1}$
$\tilde{\xi}(f,\tau)=\alpha\tilde{\gamma}(f,\tau-1)G^2(f,\tau)+(1-\alpha)\max[\gamma(f,\tau)-1,0]$: Priori SNR
$\tilde{\gamma}(f,\tau)=(Y_{\mathrm{R}}^2+Y_{\mathrm{I}}^2)/P_{\tilde{N}}(f)$: Posteriori SNR

Problems of generalized MMSE-STSA estimator

➢ To calculate $\tilde{\gamma}(f,\tau)$, dynamic estimation of $P_{\tilde{N}}(f)$ is required if the interference signal is nonstationary.
➢ Estimation of the shape parameter ρ , which depends on the type of target signal is required.

3. Proposed method

- We propose the use of SNMF as the interference signal estimator and estimate the shape parameter ρ using higher-order statistics.



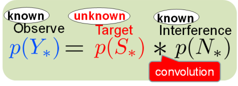
- Regarding the chi distribution, shape parameter ρ can be written using kurtosis.

kurtosis and shape parameter

$$\rho=(\mathrm{kurt}_{\mathrm{target}}-1)^{-1}$$

$\mathrm{kurt}_{\mathrm{target}}=\mu_4/\mu_2^2$: Kurtosis of target signal
$\mu_m=\int_{-\infty}^{\infty}x^m p(x)dx$: mth-order moment
$p(x)$: p.d.f. of target signal amplitude spectrogram

➢ Generally, p.d.f. of the additive signal is not additive.



- Shape parameter of target signal can be estimated from kurtosis of target signal amplitude spectrogram
- However, separation of statistics of additive signals are difficult.

Strategy

To cope with the mathematical problem, we introduce the cumulant.

What is the cumulant?
➢ Cumulant κ_m is the statistic which can be convert uniquely from moment.

$$\kappa_m(x)=f(\mu_1(x),\mu_2(x),...,\mu_m(x))$$
$$\mu_m(x)=g(\kappa_1(x),\kappa_2(x),...,\kappa_m(x))$$

➢ Cumulant holds the **additivity**

$$\kappa_m(Y_*)=\kappa_m(S_*+N_*)=\kappa_m(S_*)+\kappa_m(N_*)$$

Convert the deconvolution of the moment into the sum of the cumulant.

Kurtosis estimation of target amplitude spectrum

- Using cumulant, we can estimate kurtosis of the target amplitude spectrum as follows.

Kurtosis of target amplitude spectrogram (complex-domain)

$$\mathrm{kurt}_{\mathrm{target}}=\frac{\mu_4((S_{\mathrm{R}}^2+S_{\mathrm{I}}^2)^{\frac{1}{2}})}{\mu_2^2((S_{\mathrm{R}}^2+S_{\mathrm{I}}^2)^{\frac{1}{2}})}=\frac{\mathcal{N}\left(\mu_m(Y_{\mathrm{R}}),\mu_m(Y_{\mathrm{I}}),\mu_m(N_{\mathrm{R}}),\mu_m(N_{\mathrm{I}})\right)}{\mathcal{D}\left(\mu_m(Y_{\mathrm{R}}),\mu_m(Y_{\mathrm{I}}),\mu_m(N_{\mathrm{R}}),\mu_m(N_{\mathrm{I}})\right)}$$

$$\begin{aligned}&\mathcal{N}\left(\mu_m(Y_{\mathrm{R}}),\mu_m(Y_{\mathrm{I}}),\mu_m(N_{\mathrm{R}}),\mu_m(N_{\mathrm{I}})\right)\\&=\mu_4(Y_{\mathrm{R}})+\mu_4(Y_{\mathrm{I}})-\mu_4(N_{\mathrm{R}})-\mu_4(N_{\mathrm{I}})\\&\quad+6\mu_2^2(N_{\mathrm{R}})+6\mu_2^2(N_{\mathrm{I}})+2\mu_2(Y_{\mathrm{R}})\mu_2(Y_{\mathrm{I}})\\&\quad+2\mu_2(N_{\mathrm{R}})\mu_2(N_{\mathrm{I}})-6\mu_2(Y_{\mathrm{R}})\mu_2(N_{\mathrm{R}})\\&\quad-6\mu_2(Y_{\mathrm{I}})\mu_2(N_{\mathrm{I}})-2\mu_2(Y_{\mathrm{R}})\mu_2(N_{\mathrm{I}})\\&\quad-2\mu_2(Y_{\mathrm{I}})\mu_2(N_{\mathrm{R}})\end{aligned}$$

$$\begin{aligned}&\mathcal{D}\left(\mu_m(Y_{\mathrm{R}}),\mu_m(Y_{\mathrm{I}}),\mu_m(N_{\mathrm{R}}),\mu_m(N_{\mathrm{I}})\right)\\&=\mu_2^2(Y_{\mathrm{R}})+\mu_2^2(Y_{\mathrm{I}})+\mu_2^2(N_{\mathrm{R}})+\mu_2^2(N_{\mathrm{I}})\\&\quad+2\mu_2(Y_{\mathrm{R}})\mu_2(Y_{\mathrm{I}})-2\mu_2(Y_{\mathrm{R}})\mu_2(N_{\mathrm{R}})\\&\quad-2\mu_2(Y_{\mathrm{R}})\mu_2(N_{\mathrm{I}})-2\mu_2(Y_{\mathrm{I}})\mu_2(N_{\mathrm{R}})\\&\quad-2\mu_2(Y_{\mathrm{I}})\mu_2(N_{\mathrm{I}})+2\mu_2(N_{\mathrm{R}})\mu_2(N_{\mathrm{I}})\end{aligned}$$

- In SNMF, **only an amplitude spectrum is obtained.**

- **Represent above formula in amplitude-spectrogram domain** assuming that the real and imaginary parts are i.i.d.

➢ Assuming the i.i.d., we obtain the following relations.

$$\mu_2(Y_{\mathrm{R}})=\mu_2(Y_{\mathrm{I}})=\frac{1}{2}\mu_2(|Y|) \qquad \mu_4(Y_{\mathrm{R}})+\mu_4(Y_{\mathrm{I}})=\mu_4\left(|Y|\right)-\frac{1}{2}\mu_2^2(|Y|)$$

$$\mu_2(N_{\mathrm{R}})=\mu_2(N_{\mathrm{I}})=\frac{1}{2}\mu_2(|N|) \qquad \mu_4(N_{\mathrm{R}})+\mu_4(N_{\mathrm{I}})=\mu_4\left(|N|\right)-\frac{1}{2}\mu_2^2(|N|)$$

$|Y|=(Y_{\mathrm{R}}^2+Y_{\mathrm{I}}^2)^{\frac{1}{2}}$: Amplitude spectrogram of target signal
$|N|=(N_{\mathrm{R}}^2+N_{\mathrm{I}}^2)^{\frac{1}{2}}$: Amplitude spectrogram of interference signal

- Using these relations, we can rewrite kurtosis estimation formula as follows

Kurtosis of target amplitude spectrogram (amplitude-domain)

$$\mathrm{kurt}_{\mathrm{target}}=\frac{\mu_4(|Y|)-\mu_4\left(|N|\right)+4\mu_2^2(|N|)-4\mu_2(|Y|)\mu_2(|N|)}{\mu_2^2(|Y|)+\mu_2^2(|N|)-2\mu_2(|Y|)\mu_2(|N|)}$$

All the estimates can be obtained from the result of SNMF without using any waveforms.
➢ $|Y|$ is obtained by observed signal.
➢ $|N|$ is obtained by SNMF output.

We can calculate kurtosis of target amplitude spectrogram in closed-form

4. . Evaluation experiment

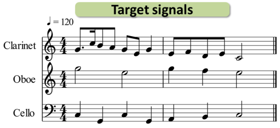
- Extract the target signal from observed signal.
- To confirm the effectiveness of the proposed method, we compared the three conventional method with our proposed method.

Experimental condition

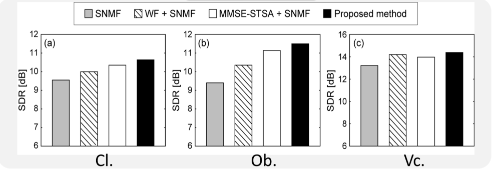
Target instruments (MIDI)	Ob., Cl., Vc.
Observed signal (MIDI)	Mixing two sources selected from three sources with the same power
Supervision sound (MIDI)	Artificial MIDI sounds of the target instruments that consists two octave notes, which cover all notes of the target signal
Compared method	SNMF Wiener filter + SNMF (WF+SNMF) MMSE-STSA estimator(Gaussian distribution)＋SNMF (MMSE-STSA+SNMF) Generalized MMSE-STSA estimator + SNMF (Proposed method)
Evaluation scores [9]	Signal to distortion ratio (SDR: quality of extracted signal),

- Wiener filter (WF) and the MMSE-STSA estimator utilized the interference spectrogram estimated by SNMF.

Target signals



Average scores



- We can confirm that the separation performance of the proposed method is better than those of the other methods.

- This result indicates the efficacy of introducing the flexible a priori statistical model of the target signal.

4. . Conclusion

- **We propose a new approach for addressing music signal separation based on the generalized Bayesian estimator with "automatic prior adaptation".**
- **From the experimental evaluation, it is found that the proposed method outperforms competitive methods, namely, simple SNMF, WF, and the MMSE-STSA estimator with a fixed Gaussian prior.**

References

[1] I. Andrianakis, P. R. White, "MMSE speech spectral amplitude estimators with chi and gamma speech priors," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 1071*, pp.III-1068–III-1071, 2006.

[2] C. Breithaupt, M. Krawczyk, R. Martin, "Parameterized MMSE spectral magnitude estimation for the enhancement of noisy speech," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing,* pp.4037–4040,2008.

[3] P. Smaragdis, B. Raj, M. Shashanka, "Supervised and semi-supervised separation of sounds from single-channel mixtures," *Proc. 7th International Conference on Independent Component Analysis and Signal Separation*, pp.414–421, 2007.

[4] K. Yagi, Y. Takahashi, H. Saruwatari, K. Shikano, K. Kondo,Music signal separation by orthogonality and maximum-distance constrained nonnegative matrix factorization with target signal information," *Proc. AES 45th Conference on Applications of Time-Frequency Processing in Audio*, 2012.

[5] D. D. Lee, H. S. Seung, "Algorithms for non-negative matrix factorization,"Proc. Advances in Neural Information Processing Systems, vol.13, pp.556–562, 2001. 2001.