

# 時間微分スペクトログラムを用いたブラインド音源分離 Blind Audio Source Separation Using Time Differential of Spectrograms

綾野 翔馬<sup>1</sup> 北村 大地<sup>1</sup> 矢田部 浩平<sup>2</sup>  
 S. Ayano<sup>1</sup> D. Kitamura<sup>1</sup> K. Yatabe<sup>2</sup>  
 (香川高等専門学校<sup>1</sup>, 東京農工大学<sup>2</sup>)

## 1. はじめに

複数の音源信号が混合した観測信号から、事前情報を用いずに混合前の音源信号を推定するブラインド音源分離 (BSS) が広く研究されている。多くの BSS 手法は、観測信号に短時間 Fourier 変換 (STFT) を施して得られる複素スペクトログラムのうち、振幅の時間周波数 (TF) 構造を手がかり (音源モデル) に用いている。一方、位相の TF 構造も考慮した音源モデルを用いる例は比較的少ない。

本研究では、位相の TF 構造もモデル化する新しい BSS を目指す。特に、瞬時周波数 (IF) に基づく修正位相スペクトログラム (iPC) [1] という TF 表現に着目し、これを BSS に用いる方法を想定する。その基礎検討として、修正位相スペクトログラムと深く関連する時間微分複素スペクトログラムを BSS に適用した結果について報告する。

## 2. 位相の TF 構造を含む音源モデルの検討

位相スペクトログラムは周期性や時間的に持続する変動等に起因して、TF 構造をモデル化することが難しい。そこで、新しい複素 TF 表現として iPC [1] が提案され、複素数の意味で低ランクな TF 構造のモデル化を達成した例がある [2]。BSS においても、iPC を活用すれば従来成しえなかった位相の TF 構造を含む音源モデルを導入できる。iPC は複素スペクトログラムの位相を IF に基づいて回転することで得られ、IF は複素スペクトログラムと時間微分複素スペクトログラムから非可逆な計算で得られる [1]。

混合信号の iPC を直接音源分離する場合、分離後の iPC を時間信号に戻す計算に、各音源信号の複素スペクトログラムと時間微分複素スペクトログラムが必要となり、これらは未知である。従って本稿ではまず、時間微分複素スペクトログラムの音源分離の実現可能性について調査する。

## 3. 提案手法

### 3.1. 時間微分複素スペクトログラムへの順変換と逆変換

STFT の計算の流れを Fig. 1 に示す。信号  $x = (x[l])_{l=0}^{L-1}$  の複素スペクトログラム  $X$  を得る STFT は、長さ  $L$  の短時間区間波形  $x_s = (x_s[l])_{l=0}^{L-1}$  に窓関数  $w = (w[l])_{l=0}^{L-1}$  を乗じて離散 Fourier 変換 (DFT) を適用する処理である。このとき、窓関数を  $w$  の (解析的な) 時間微分  $w' = (w'[l])_{l=0}^{L-1}$  に置き換えて STFT を適用すると、 $x$  の時間微分複素スペクトログラム  $X'$  が得られる [1]。

$X'$  を BSS する場合、分離信号の推定時間微分複素スペクトログラム  $Y'_i$  を時間波形  $y_i$  に戻す逆 STFT が必要であ

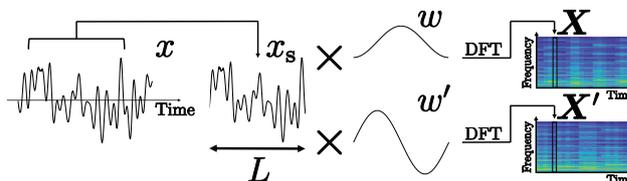


Fig. 1 Calculation procedure of STFT.

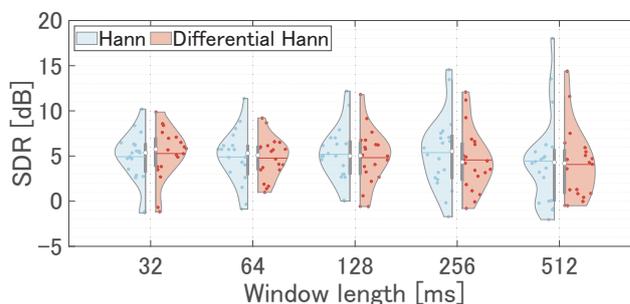


Fig. 2 BSS performance obtained by IVA.

る ( $i$  は音源番号)。そのような逆 STFT は、 $X'$  を計算する際の STFT の窓関数  $w'$  やシフト長  $S$  が次の完全再構成条件を満たす必要がある (但し、 $L/S$  は整数とする)。

$$\sum_{n=0}^{L/S-1} |w'[l+nS]| \neq 0 \quad \forall l = 0, 1, \dots, S-1 \quad (1)$$

式 (1) は、信号  $x$  のどの離散時間サンプルの情報も、 $w'$  の乗算によって失われることはないという条件である。一般的な条件の「 $w$  が Hann 窓かつ  $S = L/2$ 」では式 (1) を満たさない為、本稿は「 $w$  が Hann 窓かつ  $S = L/4$ 」を用いる。

### 3.2. 実験と結果

独立ベクトル分析 (IVA) を用いた混合音声の BSS を行った。20 個のデータセットに対して BSS を行い、得られる分離信号の信号対歪み比 (SDR) を Fig. 2 に示した。

時間微分複素スペクトログラムを用いた IVA は通常の複素スペクトログラムを用いた場合と同等の性能を示した。即ち、iPC を BSS する場合に未知となる時間微分複素スペクトログラムは、通常の BSS と同程度の精度で推定できる為、iPC を BSS に活用できる可能性が高まった。

### 参考文献

[1] K. Yatabe, et al., “Representation of complex spectrogram via phase conversion,” *Acoustical Science and Technology*, vol. 40, no. 3, pp. 170–177, 2019.  
 [2] Y. Masuyama, et al., “Low-rankness of complex-valued spectrogram and its application to phase-aware audio processing,” in *Proc. ICASSP*, pp. 855–859, 2019.