

# 音楽音響信号を対象としたドラムスの音源同定における テンプレート適応手法

吉井 和佳<sup>†</sup>      後藤 真孝<sup>‡</sup>      奥乃 博<sup>†</sup>  
<sup>†</sup> 京都大学大学院 情報学研究科 知能情報学専攻      <sup>‡</sup> 産業技術総合研究所

## 1. はじめに

ポピュラー音楽など現代の楽曲においてドラムスの果たす音楽的役割は大きいので、ドラムスの音源同定は音楽情報処理において重要な課題である<sup>1),2)</sup>。実世界の音楽音響信号を対象としたドラムスの音源同定を行う上での問題点は、

- 楽音と比べてドラム音データベースが未整備
- ドラム音は個体差が大きいので、あらゆる種類のドラム音をデータベースとして事前に準備するのは困難

である。従来研究は主にドラム単音の音源同定を対象としており、これらの問題には対処していない。

本稿では上記の問題を解決するために、テンプレート適応手法を提案する。本手法は、テンプレートとなるパワー分布を各ドラムごとに1つしか必要とせず、それを解析対象となる楽曲に使用されているドラム音に適応させていく。

## 2. テンプレート適応

本稿では、認識対象となるドラムを表現するパワー分布をテンプレート  $P_T$ 、解析対象の音楽音響信号中から抜き出したパワー分布をパターン  $P_P$  と呼ぶ。システム構成を図1に示す。テンプレート適応は、音楽音響信号中からテンプレートに近いパターンを複数抜き出し、重み付き平均を取ることによって、次世代のテンプレートを得ることで進行する。

まず、音楽音響信号から後藤らの手法<sup>2)</sup>を用いて発音時刻を粗探索しておく。この処理により、パターン抽出や選択の際の計算量を削減できる。次に、初期テンプレート(種テンプレートと呼ぶ)を生成し、パターン選択とテンプレート更新を交互に繰り返すことでテンプレート適応を実現する。

### 2.1 種テンプレート生成

$P_T$ に初期テンプレートを与えるには、ドラムの単独音が含まれた音響信号に対して以下の操作を行う。

$$P_T = \text{getPowDist}(\text{onset})$$

ここで  $\text{getPowDist}$  とは、検出された発音時刻  $\text{onset}$  から、一定時間長のパワー分布を切り出す操作である。パワー分布の計算には、窓幅 4096 点 (93ms)、窓シフト長 220 点 (5ms) の STFT を用いる。種テンプレート  $P_T$  は、時間方向に 40 frame、周波数方向に 2048 bin の大きさを持つ 2 次元配列であり、各要素は  $P_T(t, f)$  で表す。

### 2.2 パターン選択

正確なパターン抽出のために、2ms の時間分解能で発音時刻を補正する処理を行う。粗探索で得られた発音時刻候補  $T$  における補正アルゴリズムを以下に示す：

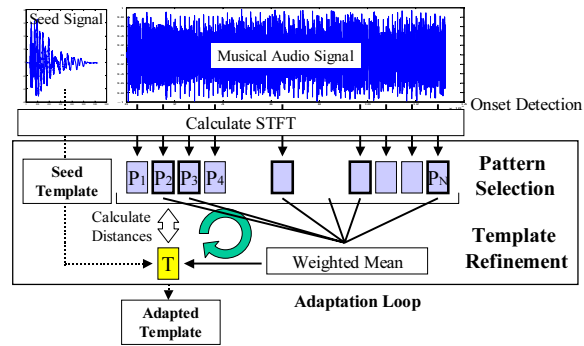


図1 Template adaptation

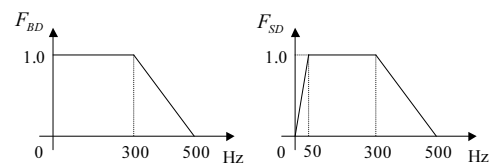


図2 Filter functions of bass drum (BD) and snare drum (SD)

- (1) 以下の計算を  $i = -5, \dots, 0, \dots, 5$  について行う。
  - (a)  $P_i$  を時刻  $T + 2i$  におけるパワー分布とする。
 
$$P_i = \text{getPowDist}(T + 2i)$$
  - (b)  $P_T$  と  $P_i$  との相関値  $C(i)$  を求める。
 
$$C(i) = \sum_{t=0}^{39} \sum_{f=0}^{2047} P_{T,w}(t, f) P_{i,w}(t, f)$$

ここで、 $P_{T,w}(t, f)$  と  $P_{i,w}(t, f)$  は次式で求める重み付きのパワー分布を表す。

$$P_{T,w}(t, f) = F_D(f) P_T(t, f)$$

$$P_{i,w}(t, f) = F_D(f) P_i(t, f)$$

ただし、 $F_D(f)$  とは、バスドラム、スネアドラムの各特性に応じたフィルター関数である(図2)。

- (2) 相関値  $C(i)$  を最大にする  $i$  を  $I$  とする。
 
$$I = \arg \max_i C(i)$$
- (3)  $P_I$  を発音時刻候補  $T$  から抽出したパターンとする。

このようにして、各発音時刻候補に対して高精度のパターンが抽出できる。得られたパターン群から、テンプレート更新に有効なパターンを複数選択するための基準として、新たに設計した改良型対数スペクトル距離尺度 (ILSDM: Improved Logarithmic Spectral Distance Measure) を用いる。

ILSDMでは、時間分解能と周波数分解能を減少させて距離を計算するために、時間軸方向、周波数軸方向ともに複数のブロックに分割する。本研究では、1つのブロックの大きさは、時間軸方向に 2 frame、周波数軸方向に 5 bin とした。抽出されたパターン総数を  $N$ 、各パターンを  $P_P^{(k)}$  ( $k = 0, \dots, N-1$ ) とする。テンプレート  $P_T$  とパターン  $P_P^{(k)}$  との間の ILSDM  $D^{(k)}$  は次式で定義する：

Template Adaptation Method on Drum Sound Identification for Musical Audio Signals by Kazuyoshi Yoshii (Kyoto Univ.), Masataka Goto (AIST), and Hiroshi G. Okuno (Kyoto Univ.)

表 1 Experimental Results

music number	method	bass drum			snare drum		
		recall rate	precision rate	F measure	recall rate	precision rate	F measure
No.6	base	89.0% (89/100)	46.4% (89/192)	0.61	95.0% (57/60)	36.8% (57/155)	0.53
	adapt	61.0% (61/100)	81.3% (61/75)	0.70	96.7% (58/60)	72.5% (58/80)	0.83
No.11	base	96.2% (51/53)	57.3% (51/89)	0.72	83.1% (54/65)	45.0% (54/120)	0.58
	adapt	96.2% (51/53)	89.5% (51/57)	0.93	73.8% (48/65)	81.4% (48/59)	0.77
No.30	base	97.3% (109/112)	94.0% (109/116)	0.96	94.0% (63/67)	63.6% (63/99)	0.76
	adapt	97.3% (109/112)	100.0% (109/109)	0.99	98.5% (66/67)	94.2% (66/70)	0.96
No.50	base	98.3% (59/60)	38.1% (59/155)	0.55	98.0% (101/103)	80.1% (101/126)	0.88
	adapt	86.7% (52/60)	72.2% (52/72)	0.79	99.0% (102/103)	80.1% (101/126)	0.89
No.52	base	97.6% (121/124)	56.8% (121/213)	0.72	95.7% (67/70)	34.9% (67/192)	0.51
	adapt	91.9% (114/124)	72.6% (114/157)	0.81	68.6% (48/70)	90.6% (48/53)	0.78
average	base	95.5% (429/449)	56.1% (429/765)	0.71	93.7% (342/365)	49.4% (342/692)	0.65
	adapt	86.2% (387/449)	82.3% (387/470)	0.84	88.2% (322/365)	83.0% (322/388)	0.86

$$D^{(k)} = \sqrt{\sum_{t_b=0}^{19} \sum_{f_b=0}^{408} (\hat{P}_T(t_b, f_b) - \hat{P}_P^{(k)}(t_b, f_b))^2}$$

ここで、 $\hat{P}_T(t_b, f_b)$  は  $(t_b, f_b)$  の位置にあるユニット内のパワー合計であり、 $\hat{P}_P^{(k)}(t_b, f_b)$  も同様に定義する：

$$\hat{P}_T(t_b, f_b) = \sum_{t=2t_b}^{2t_b+1} \sum_{f=5f_b}^{5f_b+4} F_D(f) P_T(t, f)$$

$$\hat{P}_P^{(k)}(t_b, f_b) = \sum_{t=2t_b}^{2t_b+1} \sum_{f=5f_b}^{5f_b+4} F_D(f) P_P^{(k)}(t, f)$$

この操作は一種のスムージングであり、周波数成分の小さな変動を吸収することができる。したがって、実世界の音楽音響信号に対しても信頼性の高い距離を求めることが可能になる。得られた  $D^{(k)}$  ( $k = 0, \dots, N-1$ ) のうち、距離の小さいものから、ある一定の割合のパターンを選択する。

### 2.3 テンプレート更新

テンプレートの更新を、選択されたパターンの重み付け平均として定式化する。このとき、各パターン  $P_P^{(k)}$  に対する重みとして、距離  $D^{(k)}$  の逆数を用いる。すなわち、パターン選択ステージで選択されたパターン群を  $P_P^{(s)}$  ( $s = 0, \dots, M-1$ ) とすると、更新後のテンプレート  $P'_T$  は以下のように計算する：

$$P'_T(t, f) = \frac{1}{\sum_{s=0}^{M-1} \frac{1}{D^{(s)}}} \sum_{s=0}^{M-1} \frac{P_P^{(s)}(t, f)}{D^{(s)}}$$

テンプレート更新に重み付け平均を利用する理由は 2 つある。第 1 の理由は、平均計算を行うことで、ドラム以外の楽器の調波成分を減衰させつつ、ドラム固有のパワー成分のみを抽出できることである。ドラム音特有のパワーピークの位置など、選択されたパターン間で共通のパワー構造は、適応ループ内で平均計算を繰り返しても保存される。

第 2 の理由は、重みを利用することで、安定した適応過程を実現できることである。全パターンに対して選択されたパターンの割合が適切でない場合でも、現在のテンプレートとの距離が大きなパターンはテンプレート更新にほとんど寄与しない。逆に言えば、現在のテンプレートとの距離が近いパターンはテンプレート更新に大きく寄与するため、1 回の更新でテンプレートが急激に変化することを防げる。

## 3. 実験

テンプレート適応手法を用いる場合 (adapt 手法) と、用いない場合 (base 手法) との認識率を比較する。テンプレートマッチング手法には後藤らの手法<sup>2)</sup> を利用した。後藤らの開発したポピュラー音楽データベース RWC-MDB-P-2001<sup>3)</sup> から、ポピュラー曲を 5 曲を選択し、開始から 1 分を切り出して評価用データとした。種テンプレート生成には、楽器音データベース RWC-MDB-I-2001<sup>4)</sup> を用いた。すべてのデータは 16bit, 44.1kHz でサンプリングされている。

音源同定実験結果を表 1 に示す。テンプレート適応により、バスドラムの認識率が 71% から 84% に改善され、エラー削減率は 45% であった。同様に、スネアドラムの認識率が 65% から 85% に改善され、エラー削減率は 60% であった。

adapt 手法により、著しい適合率の改善が得られ、再現率の低下はわずかであった。base 手法は、再現率は非常に高いが、適合率が非常に低い。再現率と適合率のバランス (高い F 値) が重要であり、adapt 手法により大きく改善できた。テンプレート適応は、多くの場合に有効だが、過学習 (overfitting) の兆候も見られた (No.6 のバスドラムや No.52 のスネアドラムなどの認識率)。これは、選択したパターン数やテンプレート更新時の重みに問題があると考えられる。

## 4. おわりに

本稿では、実世界の音楽音響信号を対象としたドラムスの音源同定のためのテンプレート適応手法を提案した。ポピュラー音楽を対象としたバスドラムとスネアドラムの認識実験において、テンプレート適応により、平均認識率が 68% から 85% に改善された。今後は、上記で述べた過学習問題への対処やシンバルの音源同定について扱っていく予定である。

謝辞 本研究は、科研費基盤 (A) 第 15200015 号, 21 世紀 COE, およびサウンド技術振興財団研究助成を受けた。

## 参考文献

- 1) P. Herrera et al., "Automatic Labeling of Unpitched Percussion Sounds", *Proc. of AES*, 2003.
- 2) 後藤真孝, 村岡洋一, "打楽器音を対象とした音源分離システム," 信学論, J77-D-II, 5, 901-911, 1994.
- 3) 後藤真孝 他, "RWC 研究用音楽データベース: ポピュラー音楽データベースと著作権切れ音楽データベース," 情処研報, MUS-42-6, 103, 35-42, 2001.
- 4) 後藤真孝 他, "RWC 研究用音楽データベース: 音楽ジャンルデータベースと楽器音データベース," 情処研報, MUS-45-4, 40, 19-26, 2002.