

テンプレート適応を利用した 実世界の音楽音響信号に対するドラムスの音源同定

吉井 和佳[†] 後藤 真孝[‡] 奥乃 博[†]

[†] 京都大学大学院 情報学研究科 知能情報学専攻 [‡] 産業技術総合研究所

yoshii@kuis.kyoto-u.ac.jp m.goto@aist.go.jp okuno@i.kyoto-u.ac.jp

あらまし 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について述べる。このような音響信号に対してドラムスの音源同定を行う上での問題点は、曲ごとにドラムスの音色が大きく異なり、対象曲に使用されているドラムスの正確なテンプレートが事前に用意できないことである。我々は、この問題を解決するために、新しいテンプレート適応手法とテンプレートマッチング手法を提案する。まず、テンプレート適応手法を用いて、各ドラムごとに1つの基本テンプレートを、対象曲中に使用されているドラム音に適応させる。次に、距離尺度を改良したテンプレートマッチング手法を用いて、ドラムスの音源同定を行う。ポピュラー音楽を対象にした音源同定実験の結果、テンプレート適応により、バスドラムとスネアドラムの平均認識率が68%から85%まで改善された。

Drum Sound Identification for Real-World Music Using Template Adaptation Method

Kazuyoshi Yoshii[†] Masataka Goto[‡] Hiroshi G. Okuno[†]

[†] Dept. of Intelligence Science and Technology, Graduate School of Infomatics, Kyoto University

[‡] National Institute of Advanced Industrial Science and Technology (AIST)

Abstract This paper describes drum sound identification for real-world polyphonic musical audio signals. The most critical problem with drum sound identification is that acoustic features of drum sounds vary with each musical piece, and thus we cannot prepare their precise templates in advance. To solve this problem, we propose new template-adaptation and template-matching methods. The former method adapts a single *base template model* prepared for each drum sound to the corresponding drum sound appeared in the target musical piece. The latter method uses the distance measure that enables the adapted templates to be matched with the corresponding sounds in a mixture of them and other instruments. Experimental results showed that the average accuracy of identifying bass and snare drums in popular music is improved from around 68% to around 85% by the template adaptation.

1. はじめに

近年、計算機によるデジタル信号処理技術の急速な進展により、マルチメディアデータの作成や編集が一般的となった。そのようなマルチメディアデータは、今や各計算機内にとどまらず、インターネットを介して、広大な情報空間の中に存在している。特に音楽データに関しては、インターネットを通じた音楽配信技術の進歩など、音楽データ管理の重要性がますます高まっている。音楽データの氾濫に対処するためには、音楽データをアーカイブしたり、検索したりする技術が不可欠であり、音源同定技術はその根幹をなしている。

なぜなら、楽曲に用いられている楽器名の情報は、音楽データを分類したり、楽曲構造をインデキシングするときに重要な手がかりとなるからである。

ドラムスはポピュラー音楽に代表される現代の楽曲において、非常に重要な役割を果たしているため、ドラムスの音源同定技術は不可欠な技術である。現在までに調波構造を持つ楽器に対する音源同定の研究は数多く行われてきた。MartinらやEronenらは、単音を対象とした音源同定手法を検討している^{1,2)}。柏野らはテンプレート適応を利用して、アンサンブル曲に対する自動採譜システムを提案している³⁾。しかし、これらの手法は調波構造の存在を仮定しており、ドラムス

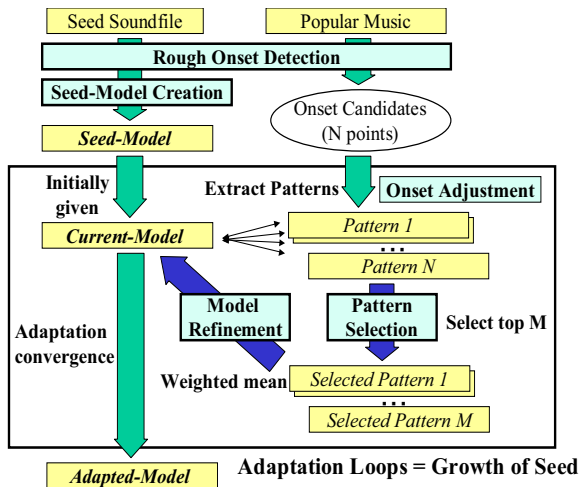


図 1 Architecture of template adaptation for a single Seed

の音源同定には適用できない。

混合音を対象としたドラムスの音源同定は十分に研究されているとはいえない。ドラムスの音源同定研究は、調波構造を持つ楽器を対象とした研究よりもずっと少なく、そのほとんどが単音を対象とした音源同定を扱ってきた。Herreraらは、643個のドラム単音を対象とした音源同定を扱い、90%程度の識別率を得ているが、混合音は対象としていない⁴⁾。Zilsらは、市販のCDトラックからのドラム音の抽出を扱っているが、対象曲中に使用されているドラム音を高音質で再合成することに焦点を当てており、音源同定については言及していない⁵⁾。実世界の音楽音響信号を対象としてドラムスの音源同定を行う上での問題点は、ドラム音は個体差に富み、各楽曲に使用されるあらゆる種類のドラム音データベースを事前に準備できないことである。

本稿では、各ドラムごとに基本テンプレートモデルを1つずつしか必要としないテンプレート適応手法を提案する。基本テンプレートモデルを1つ与えるだけで、対象曲中から対応するドラムが発音していると推定される場所を複数探索し、それら周辺のパワー分布を手がかりとして、テンプレートモデルの更新を行う処理を繰り返すことで、テンプレート適応を実現する。適応後のテンプレートモデルを用いて、距離尺度を改良したテンプレートマッチングにより音源を同定する。また、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システムを開発したので報告する。

以下、2章で本研究で提案するドラムスの音源同定システムの構成を述べる。3章でテンプレート適応手法、4章でテンプレートマッチング手法を述べ、5章でCD録音されたポピュラー音楽を対象としたドラムスの音源同定実験結果について報告する。最後に6章でまとめとする。

2. システム構成

本章では、本研究で用いる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システムの構成について述べる。音源同定システムは、主にテンプレート適応部とテンプレートマッチング部の2つの部分から構成されている。

テンプレート適応部では、各ドラムごとに用意された基本テンプレートモデルを、対象曲に使用されているドラム音をうまく表現するモデルへと適応させる。本研究では、基本テンプレートモデルのことを、“種モデル”と呼ぶことにする。1つの種（基本テンプレートモデル）を初期モデルとして与えれば、適応を繰り返すことで、よりよいモデルに自動的に成長させることが可能である。図1にテンプレート適応部の流れを示す。適応過程は、パターン選択とモデル更新の2つのステージから構成され、これらが交互に繰り返される。以下に、テンプレート適応部での各処理について簡単に述べる。

- (1) 種モデルを作るために、ドラム単音が収録されたファイルから発音時刻を検出し、そこから一定時間長のパワー分布を切り出す。
- (2) 対象曲から発音時刻を粗探索して、複数の発音時刻候補をあらかじめ求めておく。各発音時刻候補を補正し、補正後の発音時刻候補からモデルと同じ時間長のパワー分布を切り出して、各発音時刻候補から抽出したパターンとする。
- (3) 後述する距離尺度にしたがって、モデルと各パターンとの距離を計算し、距離の近いものから複数個のパターンを選択する。
- (4) 選択されたパターンの重み付き平均を計算し、次の適応ループにおけるモデルとする。

上記(2)及び(3)がパターン選択ステージ、(4)がモデル更新ステージである。

テンプレートマッチング部では、適応後のモデルと各パターンとの距離を計算し、閾値処理を行うことで、そのパターンにモデルが含まれているかを判定する。このとき、識別したいドラム以外の音が多数含まれている場合でも距離が正しく計算されるように、新しい距離尺度を提案する。この距離尺度は、時間-周波数領域において、モデルの特徴的な点の周辺に着目して距離を計算するので、他の楽器の周波数成分が混じっていたり、周波数成分の時間的なゆれがあっても、正しく距離を計算することを狙っている。

テンプレート適応部、テンプレートマッチング部についての詳細なアルゴリズムは、それぞれ3章、及び4章で述べる。

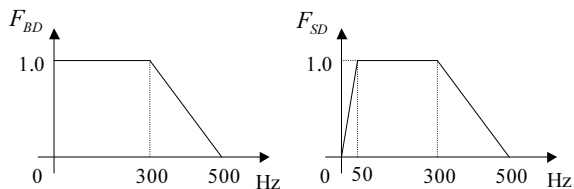


図2 Filter functions of bass drum (BD) and snare drum (SD)

3. テンプレート適応

本章では、基本テンプレートモデルである“種モデル”の適応手法について述べる。処理の流れについては2章で示した通りである。発音時刻の粗探索ステージと種モデル作成ステージは、適応ループに入るための前処理として存在する。適応ループは、パターン選択とモデル更新の2つのステージが交互に繰り返されることで実現される。

3.1 発音時刻の粗探索

対象曲中のドラムスの発音時刻を予め粗探索しておくことで、適応ループ内で行われるパターン抽出の計算量を減らすことができる。ここで、時刻 t 、周波数 f におけるパワーを $P(t, f)$ 、パワーの立ち上がり度を $Q(t, f)$ を表すことにする。 $P(t, f)$ は窓幅 4096 点、窓シフト長 10[ms] で STFT を計算することで求まる。以下に発音時刻の粗探索のアルゴリズムを示す。

- (1) $\partial P(t, f)/\partial t > 0$ が時刻 $t = a - 1, a, a + 1$ (a は任意の値) で連続して成立するとき、 $Q(a, f)$ を以下のように定める：

$$Q(a, f) = \left. \frac{\partial P(t, f)}{\partial t} \right|_{t=a}$$

そうでないなら、 $Q(a, f) = 0$ とする。

- (2) 各時刻 t について、 $S_{ins}(t)$ を、 $Q(t, f)$ の周波数 f についての重み付き和として定義する：

$$S_{ins}(t) = \sum_{f=0}^{2047} F_{ins}(f) Q(t, f) \quad (INS = BD \text{ or } SD)$$

ここで、 $F_{ins}(f)$ は、図2で示したような各ドラムの特性に応じたフィルター関数である。

- (3) Savitzky と Golay の平滑化微分の方法を用いて、 $S_{ins}(t)$ を時間 t について微分し、 $S_{ins}(t)$ が極大値をとる時刻を発音時刻候補として検出する。すなわち、 $S_{ins}(t)$ の2次微分値が正から負へ変化する時刻を検出する。

このアルゴリズムを解析対象の楽曲に適用することで、多数の発音時刻候補が求まる。この中には実際には楽器が発音していない時刻も多く含まれているので、テンプレートマッチング部の処理により、各発音時刻候補について発音しているかいないかを判定する。

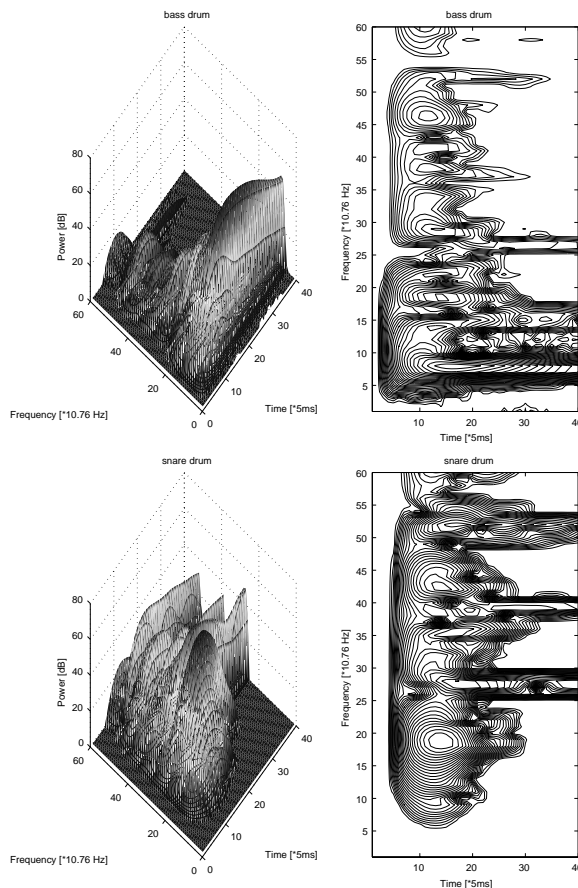


図3 Seed-Models of bass drum and snare drum

3.2 種モデル作成

種モデルは、適応ループに与えられる初期モデルである。ドラム単音が収録されたサウンドファイルから、3.1節のアルゴリズムにより検出された発音時刻を用いて、種モデルは以下のように求める：

$$P_{model}(t, f) = \text{ExtractPattern}(\text{onset})$$

ここで、 ExtractPattern とは、検出された発音時刻 onset から、一定時間長のパワー分布を切り出す操作である。このときパワー分布は、窓幅 4096 点、窓シフト長 5[ms] の STFT を計算することで求める。種モデル $P_{model}(t, f)$ は2次元配列であり、時間方向に 40[frame]、周波数方向に 2048[bin] の大きさを持つ。図3にバスドラム、スネアドラムそれぞれの種モデルの例を示す。

3.3 パターン選択

発音時刻の粗探索部での時間分解能 10[ms] は、より精度のよいパターンを抽出するために十分ではないので、正確な発音時刻を求めるために、時間分解能をあげて発音時刻補正を行う必要がある。本研究では、粗探索により求められた発音時刻の前後 10[ms] の区間を 2[ms] の時間分解能で、発音時刻を補正する処理を行う。ある発音時刻候補 T についての補正アルゴリズムは以下のように記述できる：

- (1) 以下の計算を $i = -5, \dots, 0, \dots, 5$ について行う .
 (a) $P_i(t, f)$ を時刻 $T + 2i$ から切り出したパワー分布とし , 以下の操作で求める .

$$P_i(t, f) = \text{ExtractPattern}(T + 2i)$$

- (b) $P_{model}(t, f)$ と $P_i(t, f)$ との相関値 $Corr(i)$ を以下の計算で求める .

$$Corr(i) = \sum_{t=0}^{39} \sum_{f=0}^{2047} P'_{model}(t, f) P'_i(t, f)$$

ここで , $P'_{model}(t, f)$ と $P'_i(t, f)$ は次式で表す .

$$P'_{model}(t, f) = F_{ins}(f) P_{model}(t, f)$$

$$P'_i(t, f) = F_{ins}(f) P_i(t, f)$$

- (2) 相関値 $Corr(i)$ を最大にする i を I とする .

$$I = \arg \max_i Corr(i)$$

- (3) $P_I(t, f)$ をパターンのパワー分布とし , $Pattern(t, f)$ で表すことにする .

上記のアルゴリズムにより , 各発音時刻候補に対して発音時刻補正が行われ , より精度のよいパターンが抽出される . こうして得られたパターンの中から , モデルの更新に有効なパターンを複数選択する必要がある . 本研究では , あるパターンがよりよい適応モデルを作り出すのにどのくらい貢献するかの評価基準として , 改良型対数スペクトル距離尺度 (ILSDM : Improved Logarithmic Spectral Distance Measure) を提案する . これは , モデルのパワー分布に近い形状のパワー分布を持つパターンが , 対象曲中で使用されるドラム音のモデルを構成するのに有用であるというヒューリスティクスに基づいている . 通常対数スペクトル距離尺度では , モデルとパターンのパワーのピーク位置が少しでもずれると , 距離が非常に大きくなるという問題がある . ILSDM では , そのような小さなずれを許容するように , 時間分解能 , 周波数分解能ともに減少させてから距離を計算する .

まず , 時間方向 , 周波数方向ともに複数のブロックに分割する . ILSDM は , モデル内のあるユニットとパターン内の対応するユニットとの対数スペクトル距離の和として定義される . 本研究では , 1つのユニットの大きさは , 時間方向に 2 ブロック (2[frame]), 周波数方向に 5 ブロック (5[bin]) とした . パターン総数を N とし , 各パターンを $Pattern_i(t, f)$ ($i = 0, \dots, N-1$) とする . $P_{model}(t, f)$ と $Pattern_i(t, f)$ との間の ILSDM は次式で定義される :

$$Dist_i = \sqrt{\sum_{t_b=0}^{19} \sum_{f_b=0}^{408} (P'_{model}(t_b, f_b) - Pattern''_i(t_b, f_b))^2}$$

ここで , $P'_{model}(t_b, f_b)$ 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 , $Pattern''(t_b, f_b)$ も同様に定義できる :

$$P'_{model}(t_b, f_b) = \sum_{t=2t_b}^{2t_b+1} \sum_{f=5f_b}^{5f_b+4} F_{ins}(f) P_{model}(t, f)$$

$$Pattern''(t_b, f_b) = \sum_{t=2t_b}^{2t_b+1} \sum_{f=5f_b}^{5f_b+4} F_{ins}(f) Pattern(t, f)$$

この操作はスムージングであるといえる . パワー分布をスムージングすることにより , 周波数成分の小さな変動を吸収しながら , 信頼性の高い距離を求めることが可能になる . このようにして求められた $Dist_i$ ($i = 0, \dots, N-1$) 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 .

3.4 モデル更新

更新されたモデルは , 選択されたパターンの重み付き平均として定義される . このとき , 各パターン $Pattern_i$ に対する重みとして , 距離 $Dist_i$ の逆数を用いる . すなわち , 更新されたモデルは以下のように計算できる .

$$P_{refine} = \frac{1}{\sum_{s=0}^{M-1} \frac{1}{Dist_s}} \sum_{s=0}^{M-1} \frac{Pattern_s(t, f)}{Dist_s}$$

ここで , $Pattern_s(t, f)$ ($s = 0, \dots, M-1$) は , パターン選択ステージで選択されたパターン群である .

モデル更新に重み付き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 , ドラム固有の周波数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に共通の構造は , 適応ループ内で平均計算を繰り返しても保存される . 一方 , 調波構造はパターン間で大きな変動があるので , 平均を計算すると調波構成成分は減衰する . したがって , 様々な楽器が使用されている楽曲から抽出したドラムのモデルを , 単音から作成したモデルに近づけることができる .

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 . すなわち , 一回のモデル更新で , モデルが急激に変化することを防げる . 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 , 現在のモデルとの距離が大きなパターンはモデル更新にほとんど寄与しない . 逆に言えば , 現在のモデルとの距離が近いパターンはモデル更新に大きく寄与するため , 更新後のモデルは現在のパターンから大幅に変化することがない .

上記のような理由から , モデルは適応ループを繰り返すことで徐々に変化し , 最終的にはほとんどの場合収束すると期待される .

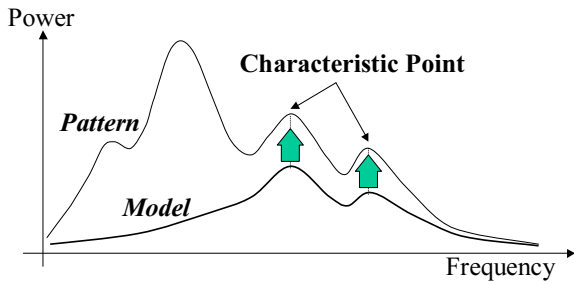


図4 Volume adjustment and matching

4. テンプレートマッチング

複数の楽器が同時に発音した場合、パターンのパワー分布にモデルが含まれていても、モデルとパターンとの距離が非常に大きくなってしまいう問題がある。この問題に対処するために、後藤らの手法⁶⁾を参考にし、新しい距離尺度を提案する。この距離尺度を用いると、あるパターンをある特定の楽器クラスに分類するのではなく、そのパターン内にモデルが含まれているかいないかを判定することが可能になる。まず、モデルとパターンとの音量補正を行った後、モデルとパターンとの距離を計算し、閾値処理により判定を行う。詳細は以下の節で述べる。

4.1 重み関数

音量補正値を決定したり、距離を計算したりする際に、パワー分布のどこに注目すればよいかを知る必要がある。そこで、時刻 t 、周波数 f ごとに注目度の大きさを表現する重み関数を利用する。

$$w(t, f) = F_{ins}(f) P_{model}(t, f) \quad (INS = BD \text{ or } SD)$$

ここで、 w は重み関数であり、 $F_{ins}(f)$ はすでに図2で示したフィルター関数である。

4.2 音量補正

モデルとパターンの音量が違くと、それらの距離が非常に大きくなるので、図4で示したように、パワー分布上の特徴点から音量補正値を計算する。音量補正アルゴリズムは以下ようになる。

- (1) 時刻 t 、周波数 f_k ($k = 0, \dots, K-1$) におけるモデルとパターンのパワーの差を $\eta(t, f_k)$ とする：

$$\eta(t, f_k) = Pattern(t, f_k) - P_{model}(t, f_k)$$

ここで f_k とは、重み関数 w の時刻 t における上位 K 点の周波数を表す。このとき、 k についての $\eta(t, f_k)$ の最小値を $\delta(t)$ と定める：

$$\delta(t) = \min_k \eta(t, f_k)$$

また、このときの $w(t, f)$ の値を $\sigma(t)$ とする。もし、 $\delta(t)$ が閾値 Θ_δ 以下のフレームの割合が、モデルのフレーム長の $R_\delta\%$ より大きいならば、パターンにモデルは含まれていないと判定する。

- (2) 音量補正値 Δ を以下の計算で求める：

$$\Delta = \frac{\sum_{\{t|\delta(t)>\Theta_\delta\}} \delta(t) \sigma(t)}{\sum_{\{t|\delta(t)>\Theta_\delta\}} \delta(t)}$$

$P_{adjust}(t, f)$ を音量補正後のモデルとすると、次式で求められる：

$$P_{adjust}(t, f) = P_{model}(t, f) + \Delta$$

4.3 距離尺度の改良

時刻 t 、周波数 f におけるモデルとパターンとの局所的な差を $\gamma(t, f)$ とする。

$$\gamma(t, f) = \begin{cases} 0 & (Pattern(t, f) - P_{adjust}(t, f) \geq -\Psi) \\ 1 & (Pattern(t, f) - P_{adjust}(t, f) < -\Psi) \end{cases}$$

ここで Ψ は正の定数である。0でない Ψ を用いることで、周波数成分のわずかな変動を吸収することができる。もし $Pattern(t, f)$ が $P_{adjust}(t, f)$ よりも大きければ、 $\gamma(t, f)$ を0としている。したがって、ドラム以外の楽器音に由来する周波数成分が含まれていても、パターンにモデルが含まれてさえいれば、局所的な距離は小さくなる。

判定に用いる距離 Γ は、 $\gamma(t, f)$ を時間-周波数領域で積分することで求まる。このとき、重み関数 w で重み付けを行う。

$$\Gamma = \sum_{t=0}^{39} \sum_{f=0}^{2047} w(t, f) \gamma(t, f)$$

ターゲットのドラムがある発音時刻候補で実際に発音しているかどうかは、 Γ に対して閾値処理することで判定できる。

5. 評価実験

本章では、ドラムスの音源同定実験について報告する。

5.1 実験条件

提案手法による認識率の改善の度合いを評価するため、実世界の音楽音響信号に対して、ドラムスの音源同定実験を行った。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) とで認識率を比較する。バスドラムとスネアドラムの認識には、それぞれ複数の閾値にすべて同じ値を用いた。

後藤らの開発したポピュラー音楽データベース RWC-MDB-P-2001⁷⁾ 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。各曲には市販CDと同様に、歌唱やドラム以外の楽器などが多く含まれている。種モデルは、同じく後藤らの開発した楽器音データベース RWC-MDB-I-2001⁸⁾ に収録されているドラム単音のサウンドファイルから作成した。すべてのデータは16bit, 44.1kHzでサンプリングされている。

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

5.2 実験結果および考察

base 手法と adapt 手法によ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。

これらの結果は、我々の提案手法が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に有効であることを示している。adapt 手法により、再現率の低下はわずかにあるものの、著しい適合率の改善が得られた。base 手法は、再現率は非常に高いが、適合率が非常に低い。我々は、再現率と適合率が高い認識率でバランスされていることが重要だと考えている。よって、adapt 手法による結果がより望ましいと言える。また、高い適合率は、テンプレート適応手法が正しく機能したことを示している。

テンプレート適応手法は、多くの場合に有効だと言えるが、しばしば、過学習 (overfitting) の兆候が見られることがある。例えば、adapt 手法による No.6 のバスドラムや No.52 のスネアドラムの再現率を見ると、base 手法に比べて大きく低下している。このようなケースでは、テンプレートモデルが少数のパターンに過剰に適応してしまったため、汎化能力に問題が生じている。この過学習問題を解決するには、モデル更新時の重み付け平均における重みを再検討したり、過学習が検出された場合、テンプレートマッチング時の閾値を適応的に変化させることが必要になる。

6. 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を実現するために、テンプレート適応手法を提案した。各ドラムの基本テンプレートモデル (種モデル) を対象曲中に実際に使用されているドラムのモデルに適応させることで、市販 CD などに現れ

る様々な個体のドラムスを扱えるようにした。我々はさらに、距離尺度を改良して、混合音中から各ドラムごとに発音時刻を検出できるテンプレートマッチング手法を開発した。

実世界の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を行った結果、テンプレート適応手法により、平均認識率が 68% から 85% に改善された。今後の課題は、適応過程におけるテンプレートモデルの過学習を解決するためのシステムの拡張である。また、シンバルや他の調波構造を持たない楽器の音源同定についても扱っていく予定である。

謝辞 本研究は、日本学術振興会科学研究費補助金基盤研究 (A) 第 15200015 号、21 世紀 COE プログラム、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) K. D. Martin, "Musical Instrument Identification: A Pattern-Recognition Approach," *136th meeting of ASA*, 1998.
- 2) A. Eronen and A. Klapuri, "Musical Instrument Recognition Using Cepstral Coefficients and Temporal Features," *ICASSP*, 753-756, 2000.
- 3) K. Kashino and H. Murase, "A Sound Source Identification System for Ensemble Music Based on Template Adaptation and Music Stream Extraction," *Speech Comm.*, **27**, 337-349, 1999.
- 4) P. Herrera *et al.*, "Automatic Classification of Drum Sounds: A Comparison of Feature Selection Methods and Classification Techniques," *ICMAI*, LNAI2445, 69-80, 2002.
- 5) A. Zils *et al.*, "Automatic Extraction of Drum Tracks from Polyphonic Music Signals," *WEDELMUSIC*, 179-183, 2002.
- 6) 後藤真孝, 村岡洋一, "打楽器音を対象にした音源分離システム," *IEICE*, **J77-D-II**, 5, 901-911, 1994.
- 7) 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," *MUS-42-6*, 103, 35-42, 2001.
- 8) 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," *MUS-45-4*, 40, 19-26, 2002.