2004－MUS－56　(8)
2004／8／3

単一テンプレート適応法による音楽音響信号を対象としたハイハットシンバルの音源同定

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

本稿では，実世界の音楽音響信号を対象としたハイハットシンバルの音源同定について扱う．打楽器の音源同定を行う上での問題点は，楽曲ごとに打楽器の音色が大きく異なり，解析対象の楽曲に含まれている打楽器音の正確なテンプレートを事前に用意できないことである．この問題を解決するため，我々はバスドラムとスネアドラムのパワースペクトルに対する単一テンプレート適応法を開発した．本稿では，ハイハットシンバル音のパワースペクトルに対する低分解能での量子化処理を導入し，単一テンプレート適応法がハイハットシンバルの音源同定にも適用可能であることを示す．ポピュラー音楽を対象にした音源同定実験の結果，単一テンプレート適応法により，ハイハットシンバルの認識精度を 48%から 82%に改善できた．

Identification of Hihat Cymbals for Musical Audio Signals Using the Single Template Adaptation Method

KAZUYOSHI YOSHII[†], MASATAKA GOTO[‡] and HIROSHI G. OKUNO[†]

[†] Dept. of Intelligence Science and Technology, Graduate School of Infomatics, Kyoto University
[‡] National Institute of Advanced Industrial Science and Technology (AIST)

This paper describes the identification of hihat cymbals for real-world polyphonic musical audio signals. The most critical problem with percussive sound identification is that acoustic features of those sounds vary with each musical piece, and thus we cannot prepare their precise sound templates in advance. To solve this problem, we developed the single template adaptation method which could be applied to power spectra of bass and snare drums. In this paper, we aim to show the effectiveness of our single template adaptation method in the identification of hihat cymbals. For this purpose, we introduce a quantization process at a lower time-frequency resolution for those power spectrum. Experimental results showed that the average accuracy of identifying hihat cymbals in popular music is improved from around 48% to around 82% by the single template adaptation method.

1. はじめに

音楽情報処理分野における重要な課題の 1 つに，コンテンツベースの楽曲検索システム[7] の実現がある．今日，計算機とインターネットの発展により，計算機上での音楽の作曲や編集は一般的になり，音楽のディジタル配信が普及している．このように楽曲の氾濫が加速する一方，ユーザの欲しい楽曲を効率的に検索する手法はいまだ実現されていない．現在の楽曲検索システムの多くがアーティスト・タイトルベースの単純な文字列検索しか行えず，音楽的なコンテンツに基づく高度な検索は研究の端緒についたばかりである．

我々は，楽曲を解析・分類するために，音楽コンテンツの 1 つとしてリズムパターン☆の側面に着目して研究に取り組む．ユーザが欲しい楽曲に対して検索要求を出すとき，アーティスト名やタイトル名に限らず，直感的なメタ情報表現（どのように音楽を知覚したか）を用いることがよくある．例えば,「ジャズ風」「ロック調」といったジャンル感に関するもの,「8 ビート」「16 ビート」といったビート感に関するもの,「ワルツ風」「スウィング的」とったリズム感に関するものなど，多様な表現が考えられる．このような人間の音楽知覚に

☆ リズムは，音楽の三大要素（メロディー，リズム，ハーモニー）のうちの 1 つである．

は，リズムパターンが密接に関係している．さまざまな音楽的な側面から楽曲を解釈し，人間の音楽知覚との関連性を明らかにすることは重要であり，第一段階としてまず，リズムパターンを取り上げる．

リズムパターンは楽曲中のドラムパートに大きく影響されるため，ドラムスの音源同定技術は不可欠である．我々は，ドラムスを構成する楽器の中でも，バスドラム・スネアドラム・ハイハットシンバルに対する音源同定手法に焦点を当てて研究を進めている．なぜなら，これら 3 種類の楽器は楽曲のテンポ，ビート[3]，拍子[6]，ジャンル[1] などの解析に応用でき，特に有用だからである．音源同定結果として得られる楽器種類と発音時刻のペアは，MPEG-7 という標準規格[2] を用いて記述することができる（**楽器タグの自動付与**）．これは，音楽コンテンツのシンボル化処理であり，高度な楽曲検索を実現するための基礎となる．MPEG-7 を用いた楽器タグの自動付与を行うことで，多種多様な楽曲に対する均質なアノテーションが期待でき，音楽コンテンツ情報の配布・再利用が容易になる．

本稿では，実世界の音楽音響信号を対象としたハイハットシンバルの音源同定について報告する．ハイハットシンバルに限らず，打楽器の音源同定を行う上での問題点は，**(A) さまざまな楽曲で使用されている打楽器の音色はバリエーションに富み，それらすべてをカバーする音テンプレートが事前に用意できないこと**，**(B) 混合音中から正しく打楽器音を認識するのが困難であること**の 2 点である．これらの問題をそれぞれ解決するために，我々はパワースペクトルに関する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開発し，バスドラム・スネアドラムの音源同定に応用した[9]．本稿では，同様のアプローチがハイハットシンバルの音源同定にも応用可能であることを示す．

テンプレート適応・マッチング手法が，バスドラム・スネアドラム・ハイハットシンバルの音源同定すべてのケースで有効に働くことは重要である．バスドラム・スネアドラムのようにスペクトル上で明確なピークを持つ音と，ハイハットシンバルのように広い周波数帯域に分布する音に対して有効に働くことは，他のさまざまな音への適用可能性が高いと考えられる．本稿では，単一テンプレート適応法にパワースペクトルに対する低分解能での量子化処理を組み込み，ハイハットシンバル音も扱えるようにする．

本稿の構成は以下の通りである．まず，2 章，3 章でテンプレート適応手法，テンプレートマッチング手法をそれぞれ説明する．次に，4 章でこれらの手法を評価するための音源同定実験について述べる．最後に，5 章でまとめとする．

2. 単一テンプレート適応法

本研究におけるハイハットシンバル音のテンプレートは，時間-周波数領域におけるパワースペクトルである．なぜなら，打楽器音のように調波構造を持たない音は，パワースペクトル形状でよく特徴付けられると考えられるからである．Zils ら[10] は，音響信号を用いてテンプレートを構成し，時間領域におけるテンプレート適応手法を提案している．本研究における単一テンプレート適応法は時間-周波数領域でテンプレートを構成するので，彼らの手法を拡張したものと見なせる．本手法をハイハットシンバルの音源同定に利用するには，1 つの「種テンプレート」を必要とする．

単一テンプレート適応法のコア部分は，適応反復アルゴリズムである．手法の概略を図 1 に示す．まず，**発音時刻粗探索ステージ**において，解析対象となる楽曲の音響信号中から発音時刻候補を粗探索しておく．そして，各発音時刻候補を開始時刻としたスペクトル断片を，楽曲のパワースペクトルから抽出する．次に，こうして抽出したすべてのスペクトル断片を用いて，種テンプレートを反復計算によって徐々に楽曲へと適応させていく．そのために，以下の 2 つのステージを，テンプレートのパワースペクトル形状が収束するまで繰り返す（反復適応アルゴリズム）．

(1) **スペクトル断片選択ステージ**　テンプレートに類似しているスペクトル断片を選択する．特別に設計した距離尺度に従い，更新前のテンプレート（種テンプレートあるいは適応反復中のテンプレート）と各スペクトル断片との距離を計算する．そして，スペクトル断片の総数に対して一定比率の個数のスペクトル断片を，距離の小さい順に選択する．

(2) **テンプレート更新ステージ**　選択したスペクトル断片の各時刻・周波数における中央値を求め，更新後のテンプレートとする．このテンプレートを，次回の適応反復における更新前のテンプレートとする．

本稿では，主にテンプレート更新ステージに改良を加えることで，ハイハットシンバルの音源同定を扱えるようにしたので報告する．

2.1 ハイハットシンバルへの対応

我々が提案した単一テンプレート適応法は，バスドラム・スネアドラムのようにスペクトル上で明確なピークを持つ音に対して有効に機能する．今回扱うハイハットシンバル音は，パワースペクトルが広い周波数帯域に分布しており，なだらかなスペクトル包絡と小さな周波数幅で大きく変動する微細なスペクトル構造を持つ．

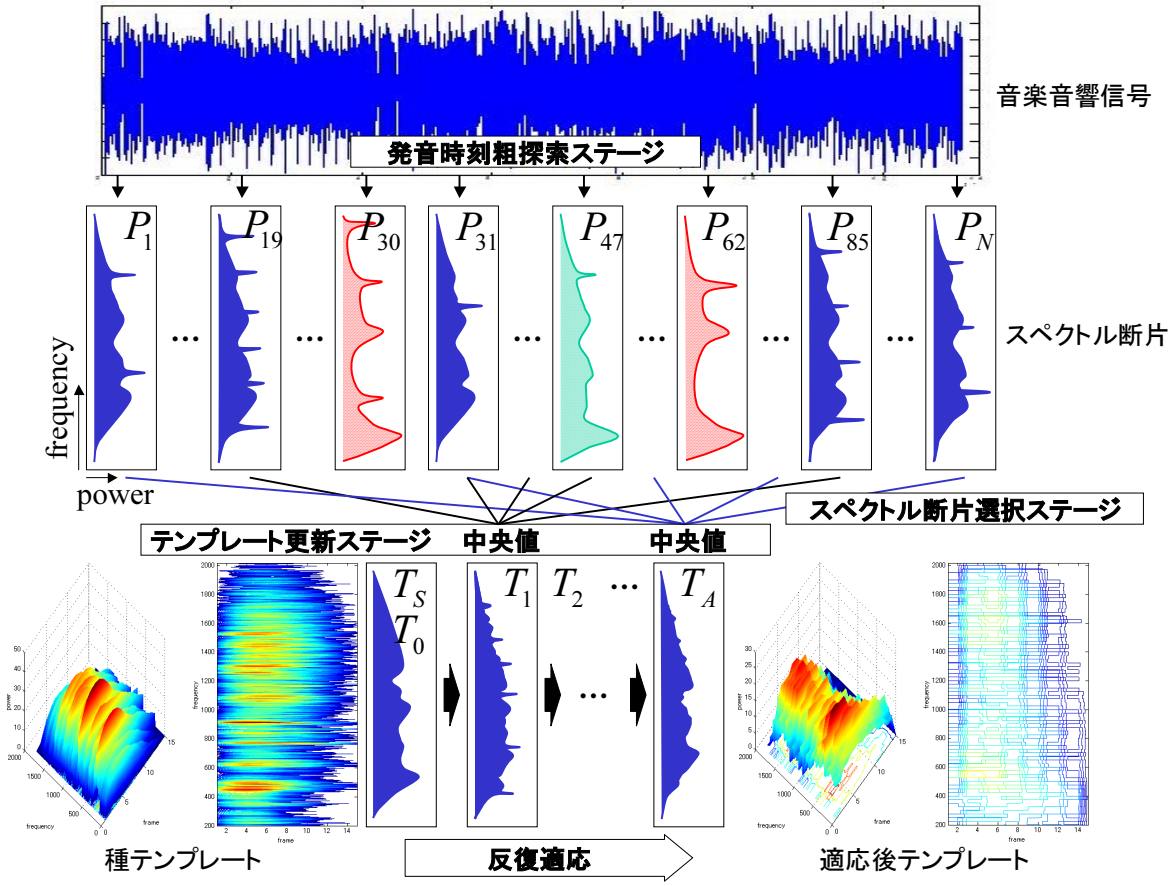

図 1　単一テンプレート適応法の概要

ハイハットシンバル音のパワースペクトルに対してテンプレート更新ステージを適用すると，更新を繰り返すたびにテンプレートの「やせ細り」が起こり，適切なテンプレートが得られない問題がある．なぜなら，ある時刻・周波数だけに着目してスペクトル断片を観察した場合，パワーの変動が激しいため，中央値を計算しても安定した値が得られないからである．

この問題の解決を解決するため，主にテンプレート更新ステージに，パワースペクトルに対する低分解能での量子化処理を導入する．パワースペクトルの分解能を下げる目的では，STFT の窓幅を小さくし，窓シフト長を大きくする方法が考えられる．しかし，シンバルのように残響が比較的長い音は，安定した発音時刻の粗探索のために窓幅を長くとり（高周波数分解能），窓シフト長を小さめ（高時間分解能）にするほうが都合がよい．また，各周波数帯域のパワー立ち上がりが発音ごとにまちまちなため，時間方向への低分解能の量子化が必要になる．本稿では，パワースペクトルを高分解能でいったん解析してから低分解能で量子化処理を行うことにする．詳細は 2.5 節で述べる．

2.2　発音時刻の粗探索

発音時刻の粗探索は，適応反復処理における 2 つのステージでの計算量を減らすために必要である．すべてのフレームからではなく，発音時刻と推測されるフレームだけからスペクトル断片を抽出することが可能になる．検出された発音時刻は，ドラムスの実際の発音時刻に必ずしも対応していない．

このステージでは，パワーの立ち上がりが十分大きいところを発音時刻候補と判断する．$P(t,f)$ をフレーム t, 周波数 f におけるパワースペクトルとし，$Q(t,f)$ を $P(t,f)$ の時間に関する微分値であるとする．$P(t,f)$ は，44.1kHz でサンプリングされた入力信号に対し，窓幅 4096 点（周波数分解能 10.8 [Hz]），窓シフト長 441 点（時間分解能 10 [ms]）のハニング窓を用いた STFT を計算することで求まる．発音時刻の粗探索のアルゴリズムを以下に示す．

(1)　時間方向に連続する 3 フレーム $t=a-1,\ a,\ a+1$ において，$\partial P(t,f)/\partial t > 0$ が満たされるとき，フレーム a における $Q(a,f)$ を以下のように定義する．

$$Q(a,f) = \left.\frac{\partial P(t,f)}{\partial t}\right|_{t=a} \qquad (1)$$

上記の条件を満たさない場合は，$Q(a,f)=0$ とする．

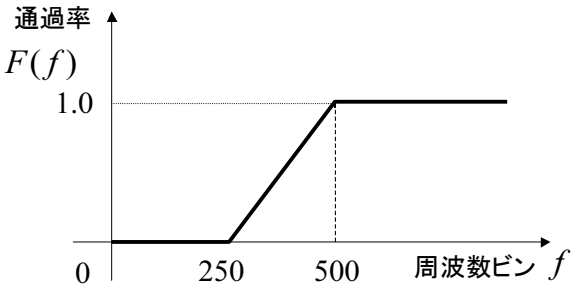

図 2 ハイハットシンバルの典型的な周波数特性を表すハイパスフィルタ関数

(2) 各フレーム t において，$S(t)$ を $Q(t,f)$ の周波数方向の重みつき和として定義する．

$$S(t) = \sum_{f=1}^{2048} F(f)\, Q(t,f) \qquad (2)$$

ここで $F(f)$ とは，図 2 で示すような，ハイハットシンバルの典型的な周波数特性を表すハイパスフィルタ関数である．これを用いることで，バスドラムやスネアドラム，歌唱などのスペクトルの影響を低減することができる．

(3) 発音時刻候補は，$S(t)$ が極大値をとる時刻として求まる．極大値を検出するには，$S(t)$ に対し Savitzky と Golay の方法[8] による 7 フレーム（前後各 3 フレーム）平滑化微分を用いる．

2.3 種テンプレート生成とスペクトル断片の抽出

単一テンプレート適応法を適用する種テンプレート T_S を生成するには，ハイハットシンバルの単音を含む音響信号が 1 つ必要になる．まず，発音時刻粗探索アルゴリズムを適用して，音響信号中の発音時刻を検出する．T_S は発音時刻を開始時刻とする一定時間長の STFT によるパワースペクトルである．T_S は行が時間，列が周波数に対応する行列であり，各要素は $T_S(t,f)$ で表す ($1 \le t \le 15$ [frames], $1 \le f \le 2048$ [bins])．適応反復アルゴリズムにおいて，g 回目の適応反復後のテンプレートを T_g とする．T_S は最初に入力されるテンプレートであるので，T_0 は T_S となる．

一方，スペクトル断片 P_i $(i = 1, \cdots, N)$ は，解析対象の楽曲中から検出された発音時刻候補 o_i [ms] を開始とする一定時間長のパワースペクトルとして抽出する．N は発音時刻候補の総数を表す．スペクトル断片 P_i はテンプレート T_g と同様の行列である．ここで，ハイパスフィルタ関数 $F(f)$ により周波数方向に重みづけられたテンプレート $\acute{T}_g$ とスペクトル断片 $\acute{P}_i$ を以下のように定義しておく．

$$\acute{T}_g(t,f) = F(f)\, T_g(t,f) \qquad (3)$$

$$\acute{P}_i(t,f) = F(f)\, P_i(t,f) \qquad (4)$$

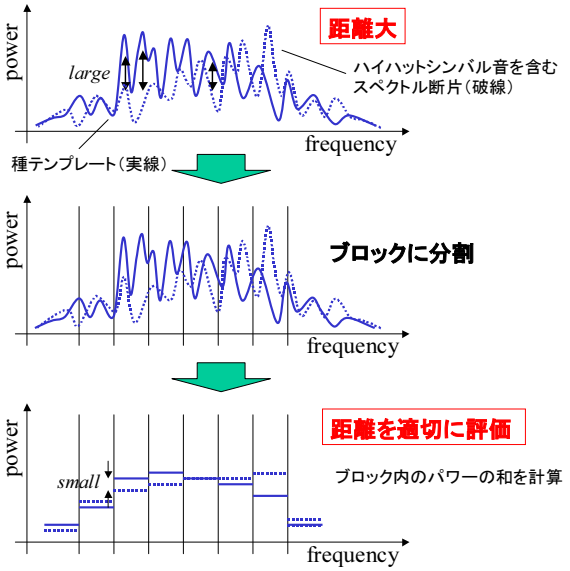

図 3 改良型対数スペクトル距離尺度の利用による効果

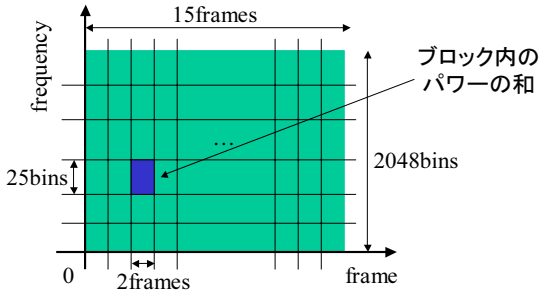

図 4 改良型の対数スペクトル尺度における低分解能での量子化処理

2.4 スペクトル断片選択

種テンプレート T_S に類似したスペクトル断片を選択するときには，図 3 に示すような改良型の対数スペクトル距離尺度を用いる．この距離尺度に従い，種テンプレートと距離が近いスペクトル断片を一定個数選択する．スペクトル断片の選択個数は，スペクトル断片の総数（発音時刻候補数）に対して一定の比率である．本稿では 0.05 とする．ここでは，通常の対数スペクトル距離尺度は利用できない．なぜなら，通常の対数スペクトル距離尺度は，スペクトルの微細構造における大きなパワー変動に敏感であるからである．すなわち，種テンプレートとスペクトル断片の音色や微細構造が少し異なるだけで，互いの距離が非常に大きくなってしまい，適切な距離計算ができなくなる．

この問題を解決するため，種テンプレートとスペクトル断片に対し，より低い時間-周波数分解能で量子化処理を行ってから距離を計算する．図 4 に概要を示すように，量子化後の時間分解能は 2 [frames] (20 [ms]) であり，周波数分解能は 25 [bins] (269 [Hz]) とする．

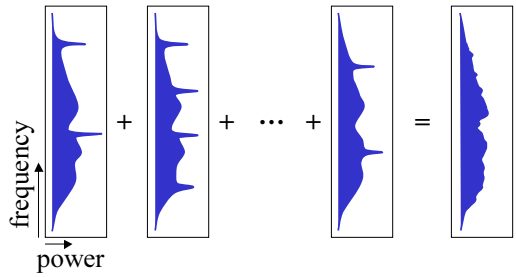

図 5　スペクトル断片の中央値によるテンプレート更新

種テンプレート $T_0(T_S)$ とスペクトル断片 P_i との間の改良型の対数スペクトル距離 $\hat{D}_i$ は次式で定義する．

$$\hat{D}_i = \sqrt{\sum_{\hat{t}=1}^{15} \sum_{\hat{f}=1}^{2048} \left(\hat{T}_0(t,f) - \hat{P}_i(t,f) \right)^2} \qquad (5)$$

ここで，低分解能での量子化後のパワースペクトル $\hat{T}_0(t,f)$ と $\hat{P}_i(t,f)$ は次式で求める．

$$\hat{T}_0(t,f) = \frac{1}{50} \left(\sum_{t'=2t-1}^{2t} \sum_{f'=25f-24}^{25f} \acute{T}_0(t',f') \right) \qquad (6)$$

$$\hat{P}_i(t,f) = \frac{1}{50} \left(\sum_{t'=2t-1}^{2t} \sum_{f'=25f-24}^{25f} \acute{P}_i(t',f') \right) \qquad (7)$$

この処理はスムージングであり，スペクトルの微細構造の違いが距離に大きく影響するのを防ぐ目的がある．まず，時間-周波数領域を 2[frames]，25[bins] の大きさのブロックに区切り，ブロック内のパワーの和を計算する．その後，もとの分解能の各ビンのパワーを計算するため，パワー和をもとのビンに再配分する．

2 回目の反復適応以降は，通常の対数スペクトル距離 D_i を利用する．

$$D_i = \sqrt{\sum_{\hat{t}=1}^{15} \sum_{\hat{f}=1}^{2048} \left(\hat{T}_g(t,f) - \acute{P}_i(t,f) \right)^2} \quad (g \geq 1) \qquad (8)$$

ここで，テンプレート更新によって得られるテンプレート $\hat{T}_g$ は，すでに低分解能で量子化処理されている（次節参照）．

2.5　テンプレート更新

T_g に対して適応処理を行い，更新されたテンプレート T_{g+1} を得るには，図 5 に示すように選択されたスペクトル断片の中央値を次式で計算する．

$$\hat{T}_{g+1}(t,f) = \underset{s}{\text{median}}\, \hat{P}_s(t,f) \qquad (9)$$

ここで，$P_s\ (s = 1, \cdots, M)$ とはスペクトル断片選択ステージで選択されたスペクトル断片である．M は選択されたスペクトル断片の個数を表す．$\hat{P}_s$ は低分解能で量子化処理されたパワースペクトルを表し，得られる更新後テンプレート $\hat{T}_{g+1}$ も同様のものになる．

$$\hat{P}_s(t,f) = \frac{1}{50} \left(\sum_{t'=2t-1}^{2t} \sum_{f'=25f-24}^{25f} \acute{P}_s(t',f') \right) \qquad (10)$$

テンプレート更新にスペクトル断片の中央値を計算する理由は，目的音以外の周波数成分を抑制するためである．ハイハットシンバル音のスペクトル構造は多数のスペクトル断片中の同じ位置に現れると期待できる．そのため，ハイハットシンバル音のスペクトル構造を持つスペクトル断片は多数派であり，中央値を計算するとその構造を抽出できる．

一方，ハイハットシンバル以外の楽器音のスペクトル成分は，選択されたスペクトル断片中の同じ位置にいつも現れるわけではない．低分解能での各フレーム・周波数における中央値を計算すると，はずれ値になりやすいそれらのスペクトル成分は抑制される．よって，ハイハットシンバル単音のテンプレートを，さまざまな楽器音を含んでいる音楽音響信号中のハイハットシンバル音に適応させることができる．

3.　テンプレートマッチング手法

本研究のテンプレートマッチング手法は，適応後のテンプレートとすべてのスペクトル断片とのマッチングを行うことで，楽曲中のハイハットシンバルの発音時刻をもれなく検出する．実世界の楽曲では，ハイハットシンバルと他の楽器とが同時発音していることがほとんどである．そのため，スペクトル断片に目的音のスペクトルが含まれていたとしても，多くの典型的な距離尺度を用いたのでは，テンプレートとスペクトル断片との距離が大きくなりすぎる．この問題を解決するため，本稿では後藤ら[5] が提案した距離尺度を改良して利用する．本稿で提案する距離尺度は，テンプレートが各スペクトル断片に含まれているかいないかに基づき距離を算出するので，他の楽器が同時発音していても正しく判定が可能である．

本手法は，適応後のテンプレート内の特徴的な時間-周波数の点に着目して距離を計算する．手法の概要を図 6 に示す．まず，**重み関数生成ステージ**において，適応後のテンプレート内の各時刻・周波数がどのくらい特徴的であるかを表す重み関数を準備する．次に，**音量補正ステージ**において，重み関数を利用して，テンプレートと各スペクトル断片との音量差が計算される．もし，音量差がある閾値よりも大きい場合は，スペクトル断片にはテンプレートは含まれていないと判定し，以降の処理は行わない．音量差があまり大きくない場合には，スペクトル断片の音量を，テンプレートの音量に合わせるように補正する．最後に，**距離計算ステージ**において，提案する距離尺度に従いテンプ

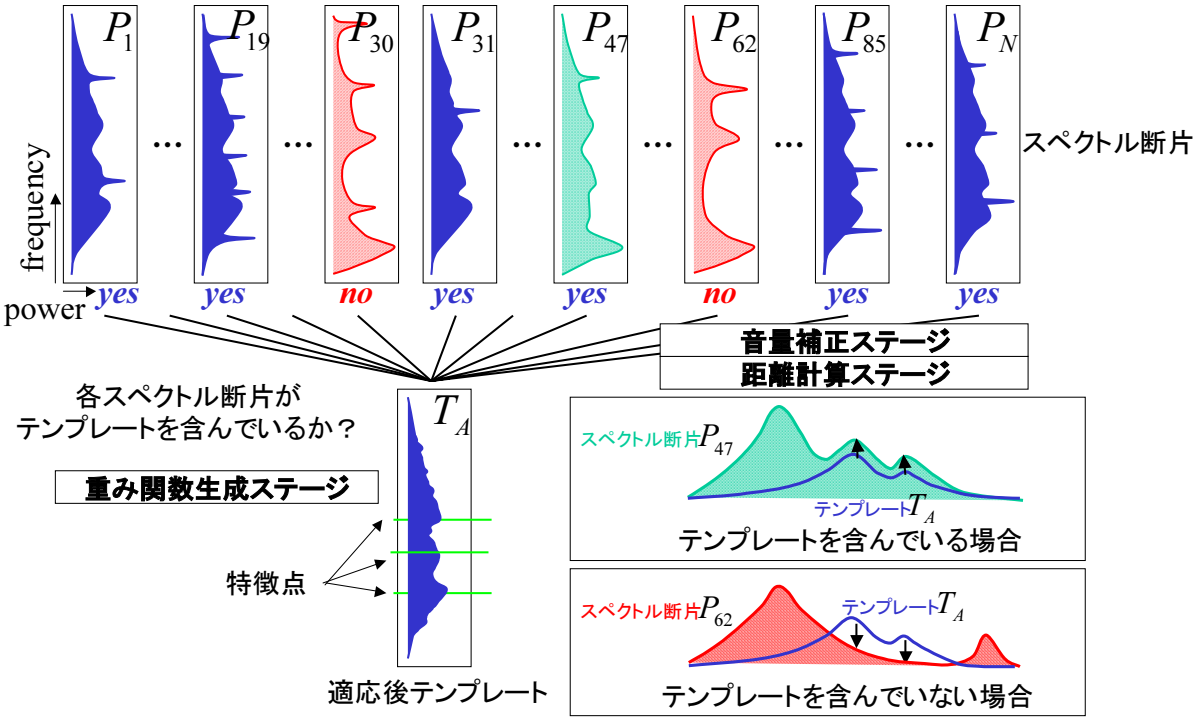

図 6　テンプレートマッチング手法の概要

レートと補正後の各スペクトル断片との距離を計算する．もし，距離がある閾値よりも小さい場合，テンプレートはスペクトル断片に含まれていると判定する．

3.1　重み関数生成

重み関数は，適応後のテンプレート内の各フレーム t，各周波数 f におけるスペクトル的な特徴の大きさを表している．重み関数 w を次式で定義する．

$$w(t,f) = \hat{T}_A(t,f) \tag{11}$$

ここで，$\hat{T}_A$ とは適応後のテンプレートであり，ハイパスフィルタ関数 $F(f)$ ですでに重みづけられている．

3.2　スペクトル断片の音量補正

適切に距離を計算するために，各スペクトル断片の音量を適応後のテンプレートの音量に合うように補正する．もし，両者の音量が異なると，テンプレートがスペクトル断片に含まれているか正しく判断できない．

テンプレート $\hat{T}_A$ とスペクトル断片 $\hat{P}_i$ との間の距離を計算するためには，$\hat{T}_A$ の行列の要素のうちでスペクトル的に特徴的な要素に着目する．まず，重み関数 w を用いて，各フレームにおける特徴点（特徴的な周波数）を求める．そして，各特徴点におけるパワーの差 η_i を計算する．次に，各フレームにおけるパワーの差 δ_i を，図 7 に示すようにそのフレームにおける η_i を用いて求める．もし，$\hat{P}_i$ のパワーが $\hat{T}_A$ よりもずっと小さい場合は，$\hat{T}_A$ は $\hat{P}_i$ には含まれていないと判定し，以降の処理は行わない（図 6 右下）．最後に，全体の音量差 Δ_i を δ_i を時間方向に積分することで求める．音量補正アルゴリズムを以下に示す．

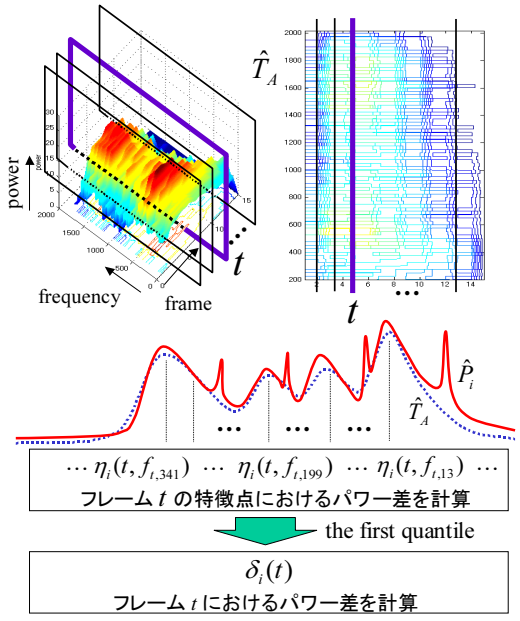

図 7　各フレーム t におけるパワー差 $\delta_i(t)$ の計算（$\eta_i(t, f_{t,k})$ の第一四分点として定義）

(1)　$f_{t,k}$ $(k = 1, \cdots, 500)$ をテンプレート $\hat{T}_A$ 中の特徴点（特徴的な周波数）とする．$f_{t,k}$ は，フレーム t において $w(t, f_{t,k})$ の値が k 番目に大きい周波数として求める．パワーの差 $\eta_i(t, f_{t,k})$ を次式で計算する．

$$\eta_i(t, f_{t,k}) = \hat{P}_i(t, f_{t,k}) - \hat{T}_A(t, f_{t,k}) \tag{12}$$

(2) フレーム t におけるパワーの差 $\delta_i(t)$ は，$\eta_i(t, f_{t,k})$ の第一四分点☆として求める．

$$\delta_i(t) = \underset{k}{\text{first-quantile}}\ \eta_i(t, f_{t,k}) \tag{13}$$

このとき，$\delta_i(t)$ をとる k の値を $K_i(t)$ とする．もし，$\delta_i(t) \geq \Psi$ を満たさないフレーム数がある閾値 R_δ よりも大きい場合，$\hat{T}_A$ は $\hat{P}_i$ には含まれていないと判定する（Ψ は負の定数である）．

(3) 最終的な音量差 Δ_i を次式で計算する．

$$\Delta_i = \frac{\sum_{\{t|\delta_i(t)>\Psi\}} \delta_i(t)\, w(t, f_{t,K_i(t)})}{\sum_{\{t|\delta_i(t)>\Psi\}} w(t, f_{t,K_i(t)})} \tag{14}$$

もし，$\Delta_i \leq \Theta_\Delta$ が満たされるなら，$\hat{T}_A$ は $\hat{P}_i$ には含まれていないと判定する（Θ_Δ はある定数）．そうでない場合，音量補正後のスペクトル断片 $\hat{P}'_i$ を次式で求める．

$$\hat{P}'_i(t, f) = \hat{P}_i(t, f) - \Delta_i \tag{15}$$

3.3 距離計算

テンプレート $\hat{T}_A$ と音量補正後のスペクトル断片 $\hat{P}'_i$ との距離を求めるには，$\hat{P}'_i$ のスペクトル中に $\hat{T}_A$ のスペクトルが含まれているか含まれていないかに着目する．もし，$\hat{P}'_i(t, f)$ が $\hat{T}_A(t, f)$ よりも大きい場合，$\hat{P}'_i(t, f)$ はハイハットシンバルのスペクトル成分だけではなくて，他の楽器のスペクトル成分が混合しているとみなす．すなわち，$\hat{T}_A(t, f)$ は $\hat{P}'_i(t, f)$ に含まれていると考える．この考え方に従い，距離尺度を次式で定義する．

$$\gamma_i(t, f) = \begin{cases} 0 & \textbf{if } \Psi \leq \hat{P}'_i(t, f) - \hat{T}_A(t, f) \leq -\Psi \\ 1 & \textbf{otherwise} \end{cases} \tag{16}$$

ここで，$\gamma_i(t, f)$ とは $\hat{T}_A$ と $\hat{P}'_i$ との間のフレーム t，周波数 f における局所的な距離である．ゼロではない負の定数 Ψ を用いることで，スペクトル成分の小さな変動を吸収する．$\hat{P}'_i(t, f)$ が $\hat{T}_A(t, f)$ 付近の値よりも大きいとき，$\gamma_i(t, f)$ は 0 になる．また，$\hat{P}'_i(t, f)$ が $\hat{T}_A(t, f)$ よりも大きすぎる場合は，$\hat{P}'_i(t, f)$ はハイハットシンバル以外の楽器のスペクトル成分がメインであると考え，$\gamma_i(t, f)$ を 1 として距離を大きくする．

全体の距離 Γ_i は，時間-周波数領域で γ_i を重み関数 w で重み付けしながら積分することで求める．

$$\Gamma_i = \sum_{t=1}^{15} \sum_{f=1}^{2048} w(t, f)\, \gamma_i(t, f) \tag{17}$$

P'_i を抽出した発音時刻でハイハットシンバルが発音したかどうかを，Γ_i をある閾値 Θ_Γ と比較することで判定する．もし，$\Gamma_i < \Theta_\Gamma$ が満たされるなら，ハイハットシンバルが発音したと判定する．

☆ 標本を小さいものから順に並べたときに，小さいものから数えて標本数の 25%の位置にあるものを第一四分点と呼ぶ．

4. 評価実験

提案手法の有効性を評価するため，実世界の音楽音響信号を対象としたハイハットシンバルの音源同定実験を行った．以下にその報告を行う．

4.1 実験条件

実験対象として，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**[4] に収録されている楽曲のうち 10 曲を用いた．各曲の最初から 1 分切り出してテストセットとした．これらには，市販 CD と同様に，ドラム音だけでなくさまざまな楽器音やボーカルが含まれている．種テンプレートは楽器音データベース **RWC-MDB-I-2001**[4] に収録されている単音のサウンドファイル 421HHCC3.WAV を用いて生成した．すべての音響信号は 16bit，44.1kHz，モノラルでサンプリングされている．

正解条件は，検出された発音時刻と実際の発音時刻とのずれが 30 [ms] 以下であることとした．また，ハイハットシンバルの奏法には主にクローズとオープンの 2 種類があり，どちらを検出しても正解とした．これらは周波数方向へのパワースペクトルの分布が似ているため識別が難しいが，発音後の残響の長さを観察することで識別できると考えられる．このような識別が必要かどうかはタスクによって異なる．

実際の発音時刻を定めるために，各楽曲の標準 MIDI ファイルからハイハットシンバルの発音時刻を抽出し，実際の発音時刻とのずれは手作業で補正した．

実験結果の評価は，再現率，適合率，F 値で行うものとし，それぞれ次式で算出する．

$$\text{再現率} = \frac{\text{正解した発音時刻数}}{\text{実際の発音時刻数}}$$

$$\text{適合率} = \frac{\text{正解した発音時刻数}}{\text{提案手法により検出された発音時刻数}}$$

$$\text{F 値} = \frac{2 \cdot \text{再現率} \cdot \text{適合率}}{\text{再現率} + \text{適合率}}$$

4.2 音源同定実験結果

テンプレート適応後にテンプレートマッチングを行う手法（adapt 手法と呼ぶ）と，テンプレート適応なしでテンプレートマッチングを行う手法（base 手法と呼ぶ）とで比較実験を行った．base 手法においてテンプレートマッチングに用いるテンプレートは適応後のテンプレートではなく，種テンプレートである．各実験で，表 1 に示すような閾値をそれぞれ用いた．

表 1 比較実験に用いる閾値

method	R_δ [frames]	Ψ [dB]	Θ_Δ [dB]	Θ_Γ
base	7	-10	-10	90000
adapt	7	-5	-4	90000

表 2 ポピュラー音楽 10 曲を対象とした音源同定実験結果

piece number	base method (baseline)			adapt method (proposed)		
	recall rate	precision rate	F measure	recall rate	precision rate	F measure
No. 6	13 % (55/436)	82 % (55/67)	0.22	79 % (345/436)	81 % (345/424)	0.80
No. 11	88 % (77/88)	97 % (77/79)	0.92	100 % (88/88)	83 % (88/106)	0.91
No. 18	97 % (177/182)	76 % (177/233)	0.85	82 % (149/182)	100 % (149/149)	0.90
No. 20	95 % (108/114)	74 % (108/145)	0.83	81 % (92/114)	85 % (92/108)	0.83
No. 30	32 % (59/184)	52 % (59/114)	0.40	98 % (181/184)	54 % (181/334)	0.70
No. 44	2 % (4/235)	44 % (4/9)	0.03	92 % (216/235)	57 % (216/278)	0.70
No. 47	23 % (41/179)	87 % (41/47)	0.36	94 % (169/179)	73 % (169/230)	0.83
No. 50	90 % (163/181)	74 % (163/221)	0.81	85 % (153/181)	94 % (153/162)	0.89
No. 52	6 % (17/271)	55 % (17/31)	0.11	99 % (267/271)	86 % (267/312)	0.92
No. 61	25 % (45/183)	43 % (45/105)	0.31	98 % (179/183)	74 % (179/241)	0.84
average	36.3 % (746/2053)	71.0 % (746/1051)	0.480	89.6 % (1839/2053)	75.2 % (1839/2444)	0.818

表 2 に実験結果を示す．実験結果から，adapt 手法の有効性が分かる．ハイハットシンバルの音源同定の F 値が 10 曲の平均で 0.480 から 0.818 に改善された．このことは，単一テンプレート適応法が音色の個体差を吸収したことを示している．また，低分解能での量子化処理により，テンプレート更新のたびにスペクトルがやせ細る現象は見られなかった．

多くの楽曲で，adapt 手法では再現率が大幅に改善された．base 手法では，ごくわずかの発音時刻しか検出できないことがしばしばあった（No. 44 や No. 52 など）．これは，種テンプレートとスペクトル断片との距離が適切に計算されなかったからである．すわなち，音色差が大きいため距離が非常に大きくなり，閾値で打ち切られてしまい発音時刻がほとんど検出できなかった．

adapt 手法により F 値を改善できたが，適合率が低いままの楽曲が少数存在する．例えば，No 30 や No. 44 における F 値の向上は，再現率の大幅な改善によるものであり，適合率はほとんど改善されていない．このような楽曲では，ハイハットシンバルの音色があまりくっきりしていないため，発音時刻の粗探索で，パワーの立ち上がりが大量に検出された．そのため，真の発音時刻は正しく検出できたが，真の発音時刻から少しずれていても発音していると判定されてしまい，適合率を下げたのが原因である．

5. おわりに

本稿では，実世界の音楽音響信号を対象としたハイハットシンバルの音源同定手法について述べた．もし，準備した種テンプレートが解析対象の楽曲で使用されているハイハットシンバル音のパワースペクトルと異なっても，単一テンプレート適応法により，種テンプレートを適応させることで対処できた．このとき，パワースペクトルに対する低分解能での量子化処理を組み込むことで，ハイハットシンバルを扱えるように手法を拡張した．市販 CD と同等のポピュラー音楽を用いた音源同定実験の結果，単一テンプレート適応法により音源同定率が大きく改善されることが示せた．

これまでの研究で，リズムパターンに密接に関係するバスドラム・スネアドラム・ハイハットシンバルの音源同定が可能になった．今後は，リズムパターンに着目した楽曲検索システムの構築を目指す．

謝辞 本研究は，科研費基盤 (A) 第 15200015 号および 21 世紀 COE の研究助成を受けた．有益なご助言を下さった駒谷和範助手，尾形哲也講師に感謝する．

参考文献

1) Dixon, S., Pampalk, E. and G. Widmer, G.: Classification of Dance Music by Periodicity Patterns, *ISMIR*, pp. 159–165 (2003).
2) Gómez, E., Gouyon, F., Herrera, P. and Amatriain, X.: Using and enhancing the current MPEG-7 standard for a music content processing tool, *AES* (2003).
3) Goto, M.: An Audio-based Real-time Beat Tracking System for Music With or Without Drum-sounds, *Journal of New Music Research*, Vol. 30, No. 2, pp. 159–171 (2001).
4) 後藤真孝, 橋口博樹, 西村拓一, 岡隆一: RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース, 情報処理学会論文誌, Vol. 45, No. 3, pp. 728–738 (2004).
5) 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, 信学論 D-II, Vol. J77-D-II, No. 5, pp. 901–911 (1994).
6) Gouyon, F. and Herrera, P.: Determination of the meter of musical audio signals: Seeking recurrences in beat segment descriptors, *AES* (2003).
7) Pampalk, E., Dixon, S. and Widmer, G.: Exploring Music Collections by Browsing Different Views, *ISMIR*, pp. 201–208 (2003).
8) Savitzky, A. and Golay, M.: Smoothing and Differentiation of Data by Simplified Least Squares Procedures, *J. of Analytical Chemistry*, Vol. 36, No. 8, pp. 1627–1639 (1964).
9) 吉井和佳, 後藤真孝, 奥乃博: 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, 情報処理学会研究報告, MUS-53-2003, pp. 55–60 (2003).
10) Zils, A., Pachet, F., Delerue, O. and Gouyon, F.: Automatic Extraction of Drum Tracks from Polyphonic Music Signals, *WEDELMUSIC*, pp. 179–183 (2002).