3R-7

調波構造の抑制によるドラムス発音時刻検出の頑健化

吉井 和佳 † 後藤 真孝 ‡ 駒谷 和範 † 尾形 哲也 † 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 (AIST)

1. はじめに

音楽情報検索システムを実現するためには，市販 CD レベルの複雑な音楽音響信号を対象とした，バスドラム音とスネアドラム音（どちらかを指す場合単にドラム音と呼ぶ）の発音時刻検出は重要な課題である．我々はこれまでに，ドラム音の音響的な特徴は個体差が大きいこと，ドラム音以外の楽器音の重畳により音響的な特徴が変動することの 2 つの問題に対処するため，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た [1]．しかし，楽曲によっては調波構造がドラム音スペクトルに重畳する影響で，正しく適応やマッチングが行えないという課題が残されていた．本稿では，音楽音響信号中の調波構造の抑制処理により頑健性をさらに向上させ，発音時刻検出精度を改善することができたので報告する．

2. ドラム音の発音時刻検出手法

まず，我々の提案したドラム音の発音時刻検出手法の概要を説明する．

2.1 解決すべき問題

市販 CD レベルの音楽音響信号中のドラム音の発音時刻を検出する上での問題点は，主に以下の 2 点である．

1. さまざまな楽曲で使用されているドラム音の音響的な特徴はバリエーションに富むため，その楽曲中のドラム音の発音時刻検出に最適なテンプレートを事前に用意するのが困難であること．
2. ドラム音に複数の楽器音が重畳しているため，単純な距離尺度を用いるテンプレートマッチングでは，音響的な特徴の変動による距離の変動に対応することが困難であること．

2.2 テンプレート適応とテンプレートマッチング

2.1 節で述べた 2 つの問題を解決するために，ドラム音の一定時間長のパワースペクトルをテンプレートとし，以下のような特徴を持つ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た．

テンプレート適応手法 事前に，適当なドラム音スペクトルを初期テンプレートとして用意する．まず，テンプレートを時間方向へずらしながら音楽音響信号の部分スペクトル（テンプレートと同じ時間長を持つスペクトル）と比較し，ドラム音スペクトルを含んでいると推定される部分スペクトルを抽出する．次に，抽出された複数の部分スペクトルを手がかりにしてテンプレートを更新する．この操作を繰り返すことで，ドラム音の発音時刻検出に最適なテンプレートを得る．

テンプレートマッチング手法 ドラム音以外の楽器音の重畳に対して頑健なスペクトル距離尺度を利用し，適応後のテンプレートと音楽音響信号の部分スペクトルとのマッチングを行う．これにより，ドラム音の発音時刻を混合音中から検出する．各ドラム音テンプレートに対し，それが部分スペクトルに含まれているかいないかの yes/no 判定を行うため，バスドラム音とスネアドラム音の同時発音を認識できる．

テンプレート適応とテンプレートマッチングは，ドラム音の発音時刻検出とスネアドラム音の発音時刻検出とで独立して行う．手法の詳細は文献 [1] を参照されたい．

このドラム音の発音時刻検出手法は，ドラム音以外の楽器音の重畳に対する頑健性を考慮して設計されている．しかし，調波構造のパワーがドラム音スペクトルのパワーに比べて大きい場合には，発音時刻検出が困難であることがあった．そのため，調波構造を抑制すれば，発音時刻検出精度を改善できると考えられる．

3. 調波構造の抑制処理

本稿では，基本周波数とその整数倍の周波数を総称して「倍音」と呼ぶ．また，それらの倍音成分の重ね合わせを「調波構造」と呼ぶ．音楽音響信号中の調波構造を抑制する手法は，(1) 基本周波数の推定，(2) 倍音成分の検証，(3) 倍音成分の抑制の 3 つのアルゴリズムから構成される．これらのアルゴリズムを利用して，音楽音響信号の各時間フレームごとに調波構造の抑制を行う．

使用する音楽音響信号は 44.1kHz サンプリング, 16bit 量子化で録音されており，スペクトル計算はハニング窓を用いた短時間フーリエ変換によって行う．窓幅は 4096 点，シフト間隔は 10 [ms]（時間フレーム単位）である．以下に各アルゴリズムについて説明する．

3.1 調波構造の基本周波数推定

まず，各時間フレームごとに基本周波数を推定する．基本周波数推定には，くし型フィルタを用いて周波数解析を行う後藤らの手法 [2] を利用する．この手法は単純であるが，大きなパワーを持つ調波構造の基本周波数を推定する目的には有効に機能する．具体的には，すべての時間フレーム t，周波数 F において，F が基本周波数である可能性 $P_{F_0}(t,F)$ を評価し，各時間フレームごとにある閾値以上の $P_{F_0}(t,F)$ を持つ周波数 F を抽出する．このとき，周波数軸のスケールには対数単位の [cent] を用いる．本稿では，f_{Hz} [Hz] から f_{cent} [cent] への変換は次式によって行う．

$$f_{\mathrm{cent}} = 1200 \times \log_2 \frac{f_{\mathrm{Hz}}}{\mathrm{REF}_{\mathrm{Hz}}}, \quad \mathrm{REF}_{\mathrm{Hz}} = 440 \times 2^{\frac{3}{12}-5} \qquad (1)$$

F0 の存在可能性 $P_{F_0}(t,F)$ は次式で計算する．

$$P_{F_0}(t,F) = \int_{-\infty}^{\infty} p(x;F) P^c(t,x) dx \qquad (2)$$

ここで，周波数 x と F の単位は [cent] であり，$P^c(t,x)$ は時間フレーム t，周波数 x における音楽音響信号の振幅スペクトルの大きさである．また，$p(x;F)$ は基本周波数 F の調波構造分布モデルを表すくし型フィルタ関数であり，次式で定義する．

$$p(x;F) = \sum_{h=1}^{H} A^{h-1} G(x;F+1200\log_2 h, W_1) \qquad (3)$$

$$G(x;m,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \qquad (4)$$

Improving Robustness of Drum Sound Onset Detection by Suppressing Harmonic Structure: *Kazuyoshi Yoshii (Kyoto Univ.), Masataka Goto (AIST), Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)*

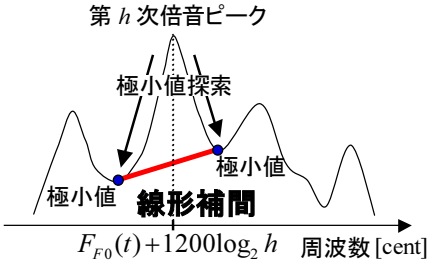

図 1: 倍音成分の抑制

ここで，H は考慮する倍音の数，A は振幅の減衰率，W_1 は倍音成分（ガウス分布 G）の裾野の広がりを意味する．本稿では，$H = 10$，$A = 0.97$，$W_1 = 150$ [cent] とした．

最後に，各時間フレーム t における基本周波数 $F_{F_0}(t)$ を，次式を満たす周波数として（複数個）求める．

$$P_{F0}(t, F_{F_0}(t)) > \max_F P_{F0}(t, F) \times \Psi_{F0} \qquad (5)$$

ここで，Ψ_{F0} は閾値であり，本稿では 0.8 とした．

3.2 倍音成分の検証

3.1 節で推定された各倍音成分が真に調波構造に由来するものであり，ドラム音スペクトルに特徴的なパワーピークと重なっていないかを検証する．3.1 節で推定された基本周波数に対応する調波構造をすべて抑制することは望ましくない．なぜなら，ドラム音のパワーが他の楽器音のパワーに比べて大きい場合，ドラム音スペクトルに特徴的なパワーピークのある周波数を，誤って基本周波数と推定する可能性があるからである．また，ドラム音スペクトルに特徴的なパワーピークがある倍音成分に重なっていた場合，ドラム音スペクトルを抑制することを防ぐ．

この判定には，一般的に，真の調波構造に由来する倍音成分はドラム音スペクトルに特徴的なパワーピークに比べて急峻なパワーピークを持つことを利用する．すなわち，各倍音のパワーピーク周辺の振幅スペクトルの 4 次モーメントを計算し，ある閾値以上の値をとる倍音成分が真の倍音成分であると判定する．本稿では，各倍音成分のパワーピークをとる周波数の上下 50 [cent] の 4 次モーメントを計算することにした．

3.3 倍音成分の抑制

3.2 節で真に調波構造に由来すると判定された倍音成分を抑制する．処理の概要を図 1 に示す．まず，周波数軸上で，倍音成分のパワーピークのすぐ両隣にあるパワー極小値をとる周波数を探索する．次に，それらの周波数の間のスペクトルを線形補間する．このとき，振幅のみを変化させ，位相は保存する．

4. 発音時刻検出実験

提案手法の有効性を評価するため，市販 CD レベルのポピュラー音楽音響信号を対象に，バスドラム音とスネアドラム音の発音時刻検出実験を行ったので報告する．

4.1 実験条件

評価対象として，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** [3] を用いた．これに収録されている楽曲には，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．初期テンプレートは楽器音データベース **RWC-MDB-I-2001** [3] に収録されている単独発音のドラム音のサウンドファイルを用いて生成した．

表 1: バスドラム音の発音時刻検出率

検出手法	再現率	適合率	F 値
TM	81.73%	79.77%	80.74%
TA + TM	90.48%	92.66%	91.56%
HS + TA + TM	92.75%	95.96%	94.34%

表 2: スネアドラム音の発音時刻検出率

検出手法	再現率	適合率	F 値
TM	50.02%	82.72%	62.34%
TA + TM	83.20%	73.87%	78.26%
HS + TA + TM	84.23%	80.81%	82.48%

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 25 [ms] 以下であることとした．実験結果の評価は，次式で定義される再現率，適合率，F 値で行う．

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

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

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

4.2 実験結果

バスドラム音とスネアドラム音の発音時刻検出実験をそれぞれポピュラー音楽 30 曲を対象として行った．各手法の検出率への寄与を評価するため，テンプレートマッチング手法 (TM)，テンプレート適応手法 (TA)，調波構造抑制手法 (HS) を 1 つずつ有効にして実験を行った．実験結果を表 1,2 に示す．実験結果から，調波構造抑制手法により平均検出率が 3.50%（エラー削減率 23.19%）向上し，手法が有効であったと言える．

テンプレート適応手法による平均検出率の向上 13.37%（エラー削減率 47.23%）ほど大きくなかったのは，我々が開発したドラム音の発音時刻検出手法が，調波構造に対する頑健性をもともと備えていたためだと考えられる．調波構造抑制手法は頑健性を補強する役割を果たし，発音時刻検出精度が改善されることが示せた．

5. おわりに

本稿では，音楽音響信号中のバスドラム音とスネアドラム音の発音時刻検出精度を向上させるために，調波構造の抑制を行う手法について述べた．具体的には，くし型フィルタを利用した周波数解析により基本周波数を推定し，その基本周波数に対応する各倍音成分を検証・抑制する．実験の結果，ドラム音の発音時刻検出精度が向上し，調波構造手法の有効性が確認できた．今後は，混合音に頑健で，より高精度な基本周波数推定手法の利用を検討していきたい．

謝辞 本研究の一部は，科学研究費補助金および 21 世紀 COE プログラムの支援を受けた．

参考文献

[1] 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 音楽情報科学研究会 研究報告 MUS-53-12, No.127, pp.55–60, 2003.

[2] M. Goto, K. Itou and S. Hayamizu: “A Real-time Filled Pause Detection System for Spontaneous Speech Recognition”, *Proc. of Eurospeech*, pp.227–230, 1999.

[3] 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.