

調波構造とHMM合成に基づく混合楽器音認識の検討*

☆山本裕貴, 山田武志, 北脇信彦, 牧野昭二 (筑波大)

1 はじめに

膨大な量の楽曲から所望の楽曲を効率良く検索する技術が求められている。検索に有用な情報の一つは楽器名であることから、個々の楽曲から楽器名を自動的に認識する混合楽器音認識の技術が重要となる。

混合楽器音認識を行う方法として、混合楽器音 HMM を用いるものがある。混合楽器音 HMM を得るための最も単純な方法は、混合楽器音データそのものを用いて学習を行うことである。しかし、楽器や音高の組合せは膨大なので現実的ではない。この問題に対して北原らは、楽器や音高の組合せを網羅していなくても性能向上に有効な学習データ集合を得られることを示した[1]。しかし、単楽器音データのみから混合楽器音 HMM を得ることができるならばより効率的であるといえる。本稿では、特徴量として調波構造を用いた場合に2つの単楽器音 HMM を合成して混合楽器音 HMM を生成する手法を提案し、その有効性を検証する。

2 提案手法

2.1 特徴量

調波構造は楽器の音色の特徴を表すことから、楽器音認識手法の多くで特徴量として用いられている。提案手法でも特徴量として調波構造を用いる。具体的には、10次倍音までを対象とし、F0 (1次倍音) のパワーと周波数 h_i (i 次倍音) のパワーの比 R_i の対数をとる。このように計算される9次元の特徴量を用いる。ここで、 $i = 2, 3, \dots, 10$ である。

楽器音は音高によって音色が変化することが知られている[2]。提案手法では、このような音色変化に対応するため、また2.3節で述べる HMM 合成の際の課題を解決するため、各楽器に対して F0 毎 (対象とする音高範囲において半音間隔毎) の HMM を学習することにした。

2.2 HMM合成

HMM 合成とは、元来音声と雑音の HMM を合成して雑音重畳音声の HMM を生成する手法として Martin らが提案したものである[3]。HMM 合成を適用することにより、2つの単楽器音 HMM を合成して混合楽器音 HMM を生成することが可能となり、混合楽器音データから直接学習する場合に比べて大幅に学習コストを削減することができる。

2.3 HMM合成の改良

提案手法では特徴量として調波構造を用いるため、ケプストラムを特徴量とするオリジナルの

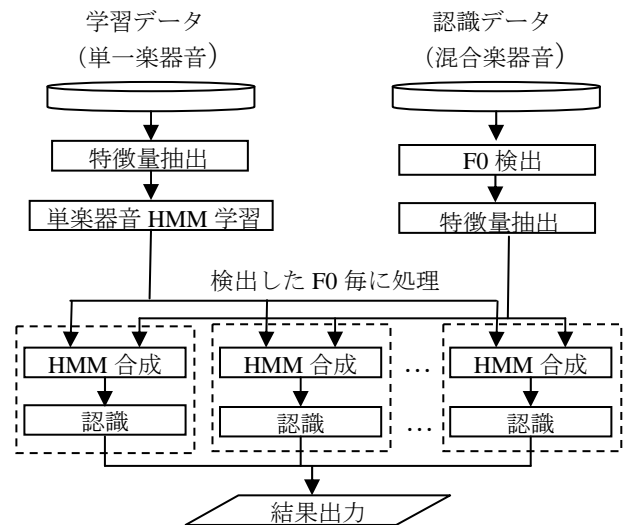


Fig.1 処理の流れ

HMM 合成をそのまま適用することはできない。

上述した h_i と R_i は F0 に依存するので、HMM 合成の改良が必要である。本稿では、以下の合成手法を提案する。

- (1) 2つの単楽器音 HMM の F0 に基づき、特徴量のどの次元に重なりが生じているのかを FFT の周波数分解能の精度で検出する。
- (2) (1)で検出された重なりのある次元のみを合成する (2.4節で詳細を述べる)。

ここで、合成後の混合楽器音 HMM の F0 は、2つの単楽器音 HMM のどちらかと一致する。F0 が一致する方の単楽器音 HMM をメイン、もう一方をサブと呼び、メインの楽器音にサブの楽器音が重畳しているとみなす。よって、提案手法では、メインとサブを入替えた2つの混合楽器音 HMM を生成し、共に認識に用いる。

2.4 合成式の定式化

メインの単楽器音 HMM を x 、サブの単楽器音 HMM を y とし、各々の特徴量の n 次と m 次に重なりが生じているとする。このとき、混合楽器音 HMM の平均 μ と分散 σ^2 は以下のように合成される。

$$\begin{aligned} \mu_{xy}(n) &= \mu_x(n) + k\mu_y(m) \\ \sigma_{xy}^2(n) &= \sigma_x^2(n) + k^2\mu_y^2(m) \\ k &= \frac{p(F0_y)}{p(F0_x)}c \end{aligned}$$

ここで、 $p(F0_x)$ はメインの F0 のパワー、 $p(F0_y)$ はサブの F0 のパワーを表す。 k は正規化のための係数、 c はメインとサブのパワー比を調整するための係数である。

* Musical instruments recognition based on harmonic structure and HMM composition, by Hiroki YAMAMOTO, Takeshi YAMADA, Nobuhiko KITAWAKI and Shoji MAKINO (University of Tsukuba).

2.5 認識の流れ

提案手法の処理の流れを Fig. 1 に示す. まず, 認識対象である混合楽器音データから各楽器音の F0 を検出し, 2.1 節で述べた特徴量を抽出する. そして, 2.3 節と 2.4 節で述べた方法により, 検出した F0 に対応する単楽器音 HMM の全組合せについて混合楽器音 HMM を生成する. これらの混合楽器音 HMM を用いて認識処理を行う. 最後に検出した F0 毎の認識結果 (混合楽器音 HMM のメインとなっている単楽器音 HMM の楽器名) を出力する. なお, 次章の実験では, 検討の初期段階として, 混合する楽器の数は 2, 各楽器の F0 は既知とする.

3 実験

3.1 実験条件

本実験では, MIDI 音源から録音した単一楽器音データ (以下では MIDI と呼ぶ) と RWC 研究用音楽データベース[4]の単一楽器音データ (以下では REAL と呼ぶ) を用いた. 対象とした楽器はピアノ, バイオリン, トランペット, クラリネット, フルートの 5 種類である. 認識用データは, 単一楽器音データを長時間平均パワーの比が 0 dB になるように混合して作成した. ここで, 楽器のペアは 25 種類である. F0 のペアは, C2 と C2~C4 (半音間隔) のペアのうち調波構造に重なりが生じる 16 種類 (以下では低音域と呼ぶ), C4 と C4~C6 (半音間隔) のペアが 9 種類である (以下では高音域と呼ぶ). なお, REAL についてはデータが不足していたために高音域のみとした.

また, 本実験では, 単楽器音 HMM の状態数は 1, 混合分布数は 1~16 とした. 提案手法として, パワー比を 0dB とした合成 HMM のみを用いた場合 (HMMC1), パワー比を -20, -10, 0, 10, 20dB とした合成 HMM を全て用いた場合 (HMMC2) の 2 種類を検討した.

3.2 実験結果

MIDI に対する認識結果を Fig. 2, REAL に対する認識結果を Fig. 3 に示す. 図中の single は単楽器音モデルのみを用いた場合, multi は混合楽器音データから直接学習した混合楽器音 HMM を用いた場合である.

図より, single の正解率は当然ながら最も低いことが分かる. 提案手法の正解率は single よりも高く, HMMC2 と HMMC1 を比較すると HMMC2 の方が優れていることが分かる. これは, 認識対象データにおける 2 つの楽器音のパワー比は短時間で見ると変動していることから, パワー比に幅を持たせている HMMC2 の方が適していることを意味する. 一方, 提案手法の正解率は multi よりも低く, 特に REAL においてはその差が顕著である. この原因については今後詳しく調査する必要がある. MIDI と REAL を比較すると, REAL の方の性能が全体的に低下している. この問題については特徴量の追加などにより対処することが考えられる.

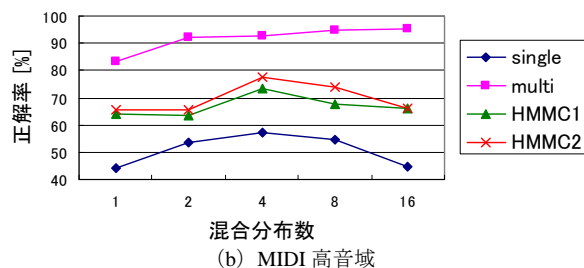
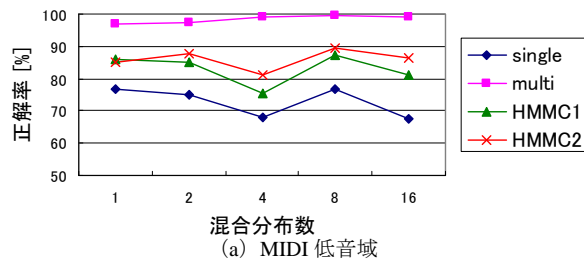


Fig.2 認識結果 (MIDI)

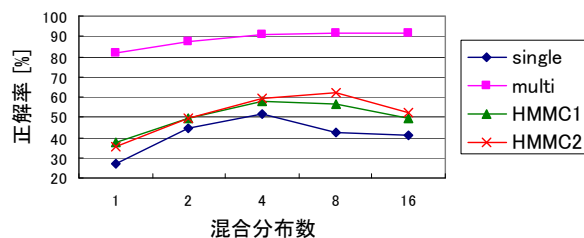


Fig.3 認識結果 (REAL)

4 おわりに

本稿では, 特徴量として調波構造を用いた場合に 2 つの単楽器音 HMM を合成する手法を提案し, 実験により一定の有効性を確認した. しかし, 混合楽器音データから直接学習した HMM を用いた場合と比べると十分な性能が得られたとは言い難い. 今後は特徴量の追加などにより性能向上を図ると共に, 同時発音数の拡大, 実演奏に対する有効性の確認を行いたい.

参考文献

- [1] T.Kitahara, *et al.*, "Instrument Identification in Polyphonic Music: Feature Weighting with Mixed Sounds, Pitch-Dependent Timbre Modeling, and Use of Musical Context," Proc. ISMIR, pp.558-563, Sept. 2005.
- [2] 北原鉄朗, *et al.*, "音高による音色変化に着目した楽器音の音源同定: F0 依存多次元正規分布に基づく識別手法," 情報処理学会論文誌, Vol.44, No. 10, pp.2448-2548, Oct. 2003.
- [3] F.Martin, *et al.*, "Recognition of noisy speech by composition of speech and noise," Proc. European Conference on Speech Communication and Technology, pp.1031-1034, 1993.
- [4] 後藤真孝, *et al.*, "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," 情報処理学会研究報告, 2002-MUS-45-4, Vol.2002, No.40, pp.19-26, May 2002.