調波構造・時間包絡・音色の統合的クラスタリング (HTTC) による複数楽器音楽信号の楽音分析*

◎宮本賢一, 亀岡弘和 †, 西本卓也, 小野順貴, 嵯峨山茂樹 (東大情報理工)

1　はじめに

本稿では、単一チャネルの複数楽器音楽信号から単音と音色のクラスタリングを同時に実現する楽音分析手法を議論する。この問題は自動楽譜作成やパート追跡、音楽検索など様々な応用が挙げられる関心の高い研究課題であるが、多重音からの音高推定や楽器・音色の分類など多くの問題が内在しており、それらを同時に分析することは極めて困難な問題とされてきた。

その要素問題の一つである多声音楽信号からの音高推定手法として、我々は HTC(Harmonic-Temporal Clustering)[1] を開発した。この手法は、従来のフレームワイズな手法 [2] とは異なり、調波構造・時間包絡を持った単音モデルを用いて、スペクトルの時間・周波数双方向の成分を単音ごとに同時にクラスタリングする方法論として高い性能を得ている。本稿ではこの HTC の発展形として、音響エネルギー成分を単音へクラスタリングしながら共通音色を有する単音の同一カテゴリへの分類していくことで、各単音の演奏情報推定と音色識別を同時に行える楽音分析の手法を検討する。

特に本研究では、人間はたとえ未知の楽器であっても異なる音色であればそれらを自然に区別し、類似する音色をグルーピングして聴くことができる点に着目し、人間のこうした教師なしの音色分類の計算論的な実現を目的とする。本稿で提案するこの統一的な楽音分析手法を Harmonic-Temporal-Timbral Clustring (HTTC) と呼ぶ。

2　問題設定

2.1　音色の定義

音色の定義やこれを決定する特徴量については、聴覚心理学の視点から様々な知見や研究がある。本研究ではこれを工学的に扱うために単純化し、音の三要素による音色の定義－大きさと高さ以外の音の性質－と、音色は音長にあまり依存しないという点に着目し、音響エネルギー領域における音色特徴量を、ピッチ・音量・発音時刻・音長に依存しない単音の音響エネルギー形状（音色形状）により定義する。

2.2　音響エネルギーの観測モデル

短時間周波数分析により、入力の音響信号から音響エネルギーを表すスペクトログラム $W(x,t)$ (x: 対数周波数、t: 時刻) が得られる。この $W(x,t)$ は、複数の音色が様々な時刻・ピッチ・音長・音量で演奏された単音の音響エネルギーの和として観測されると考えられる (図 1 参照)。よって本研究の問題は、この観測モデルの逆問題として、観測された音響エネルギーを単音のまとまり (以後音響オブジェクトと呼ぶ) にクラスタリングし、音色の定義に基づいた音響オブジェクトモデルとのフィッティングによって、各音の演奏情報を推定する問題、音色特徴量から音響オブジェクトを音色ごとに分類する問題、各音色カテゴリの音色形状を推定する問題、のすべてを同時に解くことといえる。

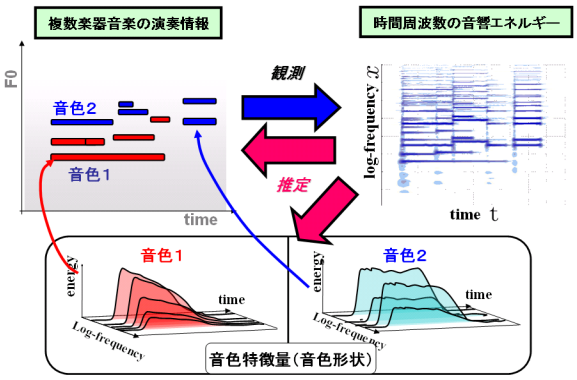

Fig. 1　音響エネルギーの観測・推定モデル

2.3　HTTC における単音・音色形状モデル

ここで単音の音響エネルギーモデル $q_k(x,t)$ について述べる。前述した音色の定義から、単音のエネルギー形状は音色カテゴリ c に依存する $T_c(x,t)$($\iint T_c(x,t)dxdt = 1$) で表現されると仮定し、各音響オブジェクトは独立したパラメータとして音量 w_k、ピッチ μ_k、発音時刻 τ_k、音長 γ_k と所属音色カテゴリ c_k を持つと定義できる (k: オブジェクト番号)。

聴覚情景解析における Bregman の分凝要件を参考に ([4])、音色形状分布 $T_c(x,t)$ を図 2 のように調波構造・連続的な時間包絡を持った形状と定義し、この分布を調波構造拘束付き 2 次元 GMM で表現する (図 2 右参照)。また前述した音色の定義より、各音において音色形状分布の時間包絡 GMM の重みに、時刻 γ_k までは 1、以降は 0 となるような連続消音分布 $R(t-\gamma_k) = \frac{1}{1+e^{p(t-\gamma_k)}}$ を積算することで、音長を考慮したエネルギーモデルが実現できる。以上の設計方針より、音響オブジェクトモデル $q_k(x,t)$ を

$$q_k(x,t) = w_k \sum_{n,y} \frac{1}{1+e^{p(y\phi-\gamma_k)}} \frac{v_{c_k,n}u_{c_k,y}}{2\pi\sigma\phi} e^{-\frac{(x-\mu_k-\log n)^2}{2\sigma^2} - \frac{(t-\tau_k-y\phi)}{2\phi^2}} \quad (1)$$

$$\sum_n v_{c,n} = \sum_y u_{c,n,y} = 1 \quad (2)$$

と表現できる。

---

* "Harmonic, Temporal and Timbral Unified Clustering for Multi-Instrumental Music Signal Analysis," by Ken-ichi Miyamoto, Hirokazu Kameoka, Takuya Nishimoto, Nobutaka Ono and Shigeki Sagayama, Graduate School of Information Schience and Technology, The University of Tokyo.

† 現在、NTT コミニュケーション基礎科学研究所に勤務

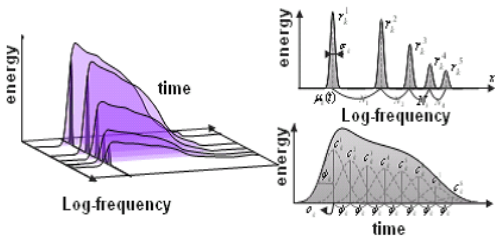

Fig. 2　音色形状分布: 概形 (左)、対数周波数・時間方向のGMM表現 (右)

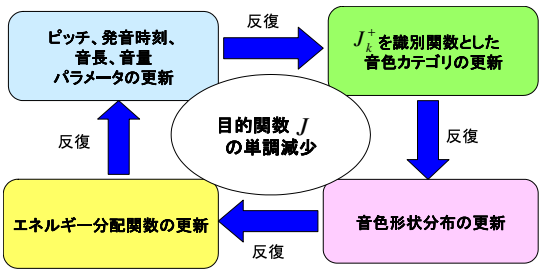

Fig. 3　目的関数最小化を実現する 4 ステップ反復アルゴリズム

3　解法: パラメータの反復推定

観測エネルギーを各音響オブジェクトに分配する分配関数 $m_k(x,t)(\sum_k m_k(x,t)=1)$ を導入し、分配されたエネルギー分布と音響オブジェクトモデル分布との近さを表す分布間距離として I ダイバージェンスを採用すると、目的関数

$$J = \sum_k \iint m_k(x,t)W(x,t)\log\left(\frac{m_k(x,t)W(x,t)}{q_k(x,t)}\right) - (m_k(x,t)W(x,t) - q_k(x,t))dxdt \quad (3)$$

を最小化する問題と定式化できる。この目的関数から、図 3 で示すような、演奏情報推定・音色カテゴリ決定・音色形状推定・エネルギーの再分配を交互に行なう 4 ステップ反復推定アルゴリズムによって、局所最適パラメータが得られる (更新式は [5] 参照)。

4　実装システムの適用例

提案アルゴリズムを実装し、実際の楽曲に適用した例を示す。入力用の楽曲は RWC 研究用音楽データベースより、ピアノとバイオリンで演奏される RWC-MDB-C-2001 No.39 の冒頭部を抜粋して利用した。本発表では音色の分類と形状学習を考察するため、MIDI 形式のデータから生成した WAV 形式の信号を入力とし、音色カテゴリ数は 2 で固定、また初期のオブジェクトの配置は MIDI を参照して実験を行なった。入力の演奏情報と音響エネルギーを図 4 に示し、推定された演奏情報と 2 つの音色形状分布を図 5 に示す。結果より、メロディ部のピアノの音色は正しく分類・推定されたが、ピアノの低音とバイオリンが同一音色カテゴリに分類される結果となった。原因としては、ピアノの倍音エネルギー比がピッチに依存していることが考えられる。

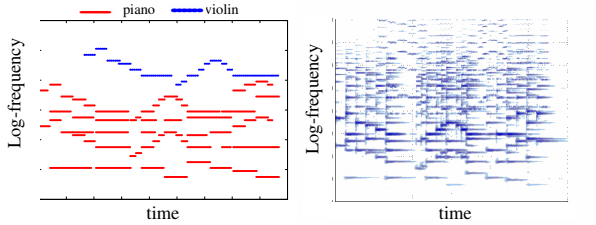

Fig. 4　左:入力楽曲の演奏情報: piano(赤), violin(青)、右: 観測音響エネルギー

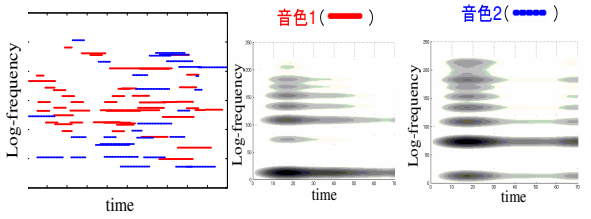

Fig. 5　推定された演奏情報と音色形状

5　おわりに

本研究では、調波構造・時間包絡の連続性・音色の類似性に基づいた音響エネルギーのクラスタリングにより、単音と音色のクラスタリングを同時に実現する楽音分析手法を提案し、また実際の楽曲を用いた適用例を示した。今後の課題として、ピッチに依存した倍音比の相違を考慮した音色モデルの設計、音響オブジェクト数の妥当性を考慮したアルゴリズムの提案、非調波構造の楽器も含めた統合的な複数音楽分析手法の提案などを検討したい。

謝辞

本研究の一部は科学研究費補助金・基盤研究 B (課題番号 17300054) および科学技術振興機構 CREST プロジェクトの補助を受けて行なわれた。

参考文献

[1] H. Kameoka *et al.* , "A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering," *IEEE Trans. on Audio, Speech and Language Processing,* in Press.

[2] M.Goto, "A Robust Predominant-F0 Estimation Method for Real-time Detection of Melody and and bass lines in CD recordings," in *Proc. ICASSP*,2000.

[3] K. Kashino, Hiroshi Murase,"A sound source identification system for ensemble music based on template adaptation and music stream extraction," *Speech Communication*,vol.27, 1999.

[4] 亀岡弘和 他, "調波時間構造化クラスタリングによる CASA へのアプローチ," 日本音響学会聴覚研究会, Vol. 36, No. 7, H-2006-103, pp. 575-580, 2006.

[5] 宮本 賢一 他, "調波構造・時間包絡・音色の統合的クラスタリングによる楽音分析," 情報処理学会研究報告, 2007-MUS-71.