Source-Filter モデルを含めた調波構造・時間包絡・音色の統合的クラスタリング (HTTC) による楽音分析*

◎宮本賢一 (東大情報理工), 亀岡弘和 (NTT 研究所),
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類似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である HTC(Harmonic-Temporal Clustering)[1] を開発しており、その発展形として、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ごとに単音を分類する統一的な楽音分析手法、Harmonic-Temporal-Timbral Clustering (HTTC)[3] を開発してきた。本研究は、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、類似する楽器音・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能力の計算論的な実現を目的とする。

特に本稿では、実際の楽器の物理モデルに即した音色特徴量として Source-Filter モデルを含めた音色モデルを提案し、その音色モデルを用いた HTTC の数理的な解法を述べ、実験により提案した音色モデルの有効性を検証する。

2 問題設定

2.1 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、後述する音色の定義に基づいた単音モデルとのフィッティングにより、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特徴量を推定する問題、のすべてを同時に解くこととして設定できる。

2.2 本稿における音色の定義

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が、我々

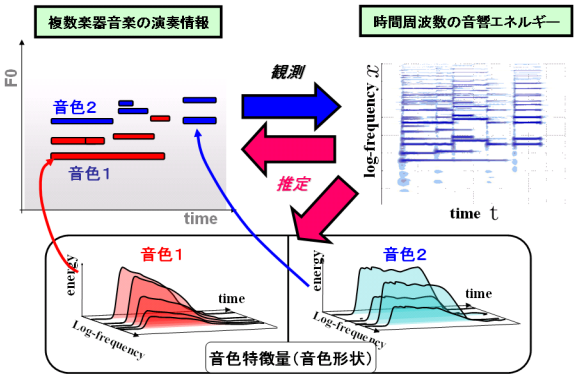

Fig. 1 音響エネルギーの観測・推定モデル

は [3] においてこれを工学的に扱うために単純化し、音響エネルギー領域における音色特徴量を、音の三要素による音色の定義に基づいてピッチ・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) と定義した。しかし実際の楽器音においては、同じ楽器であってもピッチによりエネルギー形状は変化しており、我々はその特徴を含めて一音色と感じることができる。そこで本稿では新たな特徴量として、多くの楽器の物理モデルとして妥当な振動特性 (Source モデル) と共振特性 (Filter モデル) の積による調波構造と時間方向の包絡から成る単音の 2 次元音響エネルギー形状として定義する。

2.3 Source-Filter モデルを含めた単音モデル

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 とオブジェクトのピッチ μ_k に依存する $T_c(x,t;\mu_k)$ 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 を持つと定義する (k: オブジェクト番号)。

音色形状のモデルについては、聴覚情景解析における Bregman の分凝要件を参考に ([2])、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する。周波数方向に関しては、調波構造を表す Source モデルを GMM で設計し、その重みパラメータ $v_{c,n}$ に GMM で設計された Filter モデルを積算することで (重み $\theta_{c,l}$)、ピッチによる滑らかな調波構造変化を表現する。時間方向に関しては、音色形状分布の時間包絡を GMM で設計し、各オブジェクトにおいてその重みパラメータ $u_{c,y}$ 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

* "Harmonic, Temporal and Timbral Unified Clustering with Source-Filter Model 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, and *NTT Communication Science Laboratories, NTT.

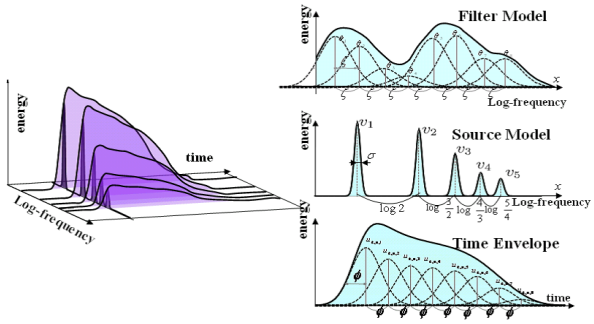

Fig. 2　音色形状分布: 概形 (左)、Source-Filter モデルと時間包絡の GMM 表現 (右)

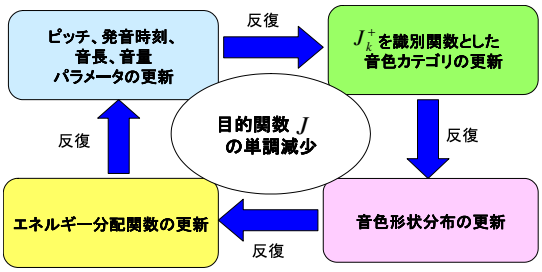

Fig. 3　目的関数最小化を実現する 4 ステップ反復アルゴリズム

方針より、音響オブジェクトモデル $q_k(x,t)$ を

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

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

と表現できる。

3　解法: パラメータの反復推定

観測音響エネルギーを各音響オブジェクトに分配する関数 $m_k(x,t)$ (但し $\sum_k m_k(x,t) = 1$) 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して定式化できる。この目的関数から、図 3 で示す 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 (更新式は省略)。

4　実音を用いた定性的評価実験

提案アルゴリズムを実装し、本稿で新たに提案した Source-Filter モデルの有効性を実音を用いて検証する。入力信号は、RWC 研究用音楽データベースのピアノとアルトサクソフォン (サックス) の単音データから、サックスの音域の 2 半音階の単音をランダム

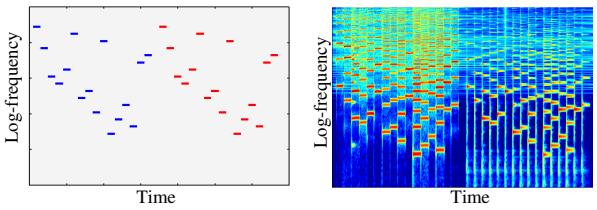

Fig. 4　左: 入力楽曲の演奏情報: alto saxophone(青), piano(赤)、右: 観測スペクトログラム

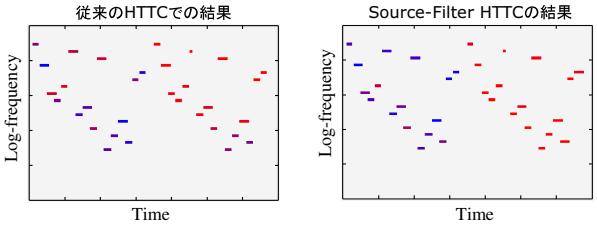

Fig. 5　音色クラスタリング結果: 左: [3] の音色モデルを用いた HTTC、右: Source-Filter モデルを含めた HTTC

に配置した信号を作成した。本稿では特に音色の分類と形状学習を考察するため、初期のオブジェクトの配置は正解を与えて実験を行なった。入力の演奏情報を図 4 に、推定結果を図 5 に示す。結果より、従来のピッチによる形状不変のモデルではピアノやサックスの低い音と高い音が別の音色に分離されたが、提案するモデルではその影響がかなり少なくなった。特にピアノは完全に一つのクラスタで表現されることを確認した。これは、本稿で提案した Source-Filter モデルによって、実際の楽器音のピッチによる形状変化をより良く表現できたからと考えられる。

5　おわりに

本研究では、Source-Filter モデルを含めた調波構造・時間包絡の連続性・音色の類似性に基づく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、特に周波数特性として Source-Filter モデルを含めることによる音色表現の妥当性を、実際の楽器音を用いた実験により検証した。今後の課題として、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, Vol. 15, No. 3, pp. 982-994, 2007.

[2] 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資料, Vol. 36, No. 7, H-2006-103, pp. 575-580, 2006.

[3] Kenichi Miyamoto, Shigeki Sagayama et al., “Harmonic-Temporal-Timbral Clustering (HTTC) for the Analysis of Multi-Instrumental Polyphonic Music Signals,” Proc. ICASSP2008, to appear.