MAP 推定を用いた Harmonic Clustering による多重音中の非調和性音源の調波構造検出*

◎織田誠也　(東大・工)　亀岡弘和　西本卓也　嵯峨山茂樹　（東大・情報理工）

1　はじめに

本報告では，多重音に含まれる非調和性楽器音の調波構造検出を扱う．発音機構や演奏方法などにより非調和性が生じる自然楽器は多く，打楽器や打弦楽器の楽器音解析や演奏分析，音源分離や音源同定など，音楽音響学や音楽情報科学の分野におけるさまざまな貢献が期待できる．

従来の多重楽器音解析の研究は，楽器音の調和性を仮定しているものが多く [1, 2]，打楽器や打弦楽器などのように必ずしも調和性をもたない楽音を厳密に扱うことはできなかった．

そこで我々が提案した Harmonic Clustering[3] を応用し，これまで固定値として設けていた基本周波数と高調波周波数との間の比例係数をモデルに新たに変数 (拘束パラメータ) として導入し，これを最大事後確率 (Maximum A Posteriori: MAP) 推定により求めることで，基本周波数だけではなく各高調波周波数の正確な推定を試みる．実際の楽器音を対象に性能評価を行い，本手法の有効性を確認した．

2　Harmonic Clustering

我々はこれまで Harmonic Clustering を提案し [3]，その一定式化として拘束つきの混合正規分布のパラメータ推定に帰着させた．本報告では，これに従って，同様な考え方で以下の議論を進める．

短時間周波数解析における観測スペクトルは，窓関数などの影響により周波数方向の広がりをもつ．この広がりが窓関数の影響のみに起因すると仮定し，正規分布窓を窓関数として用いれば，この広がりの形状は理論的に正規分布の形状と同等と見なせる．従って，単一音の調波構造スペクトルは，拘束つきの混合正規分布によりモデル化できる．これを調波構造モデルと呼ぶ．音 k の基本周波数に対応する正規分布の平均を μ_1^k とし，n 次高調波周波数の基本周波数に対する比例係数を h_n^k とすると，n 次高調波周波数に対応する正規分布の平均 μ_n^k は，$h_n^k\mu_1^k$ と表せる (例えば，調和性楽器音の場合は $h_n^k = n$ である)．

K 個の音の調波構造が重なり合う多重音スペクトルは，調波構造モデルを K 個混合することによりモデル化でき，以下のパラメータにより構成される．

$$\{\theta\} = \{\mu_1^k, w_n^k, \sigma_n^k \mid n, k \in \mathbb{N}\} \tag{1}$$

ただし，w_n^k，σ_n^k は，n 次成分の重み，分散を表す．スペクトル分布を正規化して確率変数 (周波数)ω の確率分布 $f(\omega)$ と見なせば，θ の事後確率 $p(\theta|f)$ を最大化する θ は，以下で与えられる．

$$\theta = \underset{\theta}{\operatorname{argmax}} \left\{ \log p(\theta) + \int_{-\infty}^{\infty} f(\omega) \log p(\omega|\theta) d\omega \right\} \tag{2}$$

これを解析的に解くことは困難であるが，EM(Expectation Maximization) アルゴリズムにより，以下の $Q(\theta,\bar{\theta})$ を計算する E ステップと $Q(\theta,\bar{\theta})$ を最大化する M ステップの反復計算により，θ の局所最適解を得ることができる．

$$Q(\theta,\bar{\theta}) = \log p(\bar{\theta}) + \sum_{k=1}^{K}\sum_{n=1}^{N_k}\int_{-\infty}^{\infty} p(n,k|\omega,\theta) f(\omega) \log p(\omega,n,k|\bar{\theta}) d\omega \tag{3}$$

$$p(n,k|\omega,\theta) = \frac{p(n,k,\omega|\theta)}{\sum_{k=1}^{K}\sum_{n=1}^{N_k} p(n,k,\omega|\theta)} \tag{4}$$

$$p(n,k,\omega|\theta) = \frac{w_n^k}{\sqrt{2\pi{\sigma_n^k}^2}} \exp\left\{ -\frac{(\omega-\mu_n^k)^2}{2{\sigma_n^k}^2} \right\} \tag{5}$$

ただし，N_k は楽器 k の正規分布の平均の数を表す．w_n^k，σ_n^k の最尤推定値は，M ステップにおいてそれぞれの変数を[1]以下に更新することで得られる．

$$\bar{w}_n^k = \int_{-\infty}^{\infty} p(n,k|\omega,\theta)\, d\omega \tag{6}$$

$$\bar{\sigma}_n^k = \sqrt{\frac{\int_{-\infty}^{\infty} p(n,k|\omega,\theta) f(\omega)(\omega-\mu_n^k)^2 d\omega}{\int_{-\infty}^{\infty} p(n,k|\omega,\theta) f(\omega) d\omega}} \tag{7}$$

3　非調和性調波構造モデルの定式化

この章では，これまでの Harmonic Clustering の手法を拡張し，h_n^k の揺らぎを許容した非調和性調波構造モデルを定式化する．

3.1　比例係数 h_n^k の誤差パラメータの導入

比例係数 h_n^k は，楽器の種類が事前に分かっていれば，理論値や経験的に知られる値として与えることができるが，同一楽器でも個体差や演奏方法等によって，多少の誤差が生じると十分考えられる．この誤差は，複雑な物理現象に起因していると考えられ，必ずしも予測できるものではない．そこで，その誤差を表すパラメータ ϵ_n^k を新たに導入し，事前分布を想定することで，この現象を単純化し，確率的に引き起こされるものとして扱う．従って μ_n^k は，$h_n^k\mu_1^k + \epsilon_n^k$ と表すことができ，μ_1^k を最尤推定，ϵ_n^k を MAP 推定により求めることで，非調和性楽器音の高調波周波数を得ることができる．M ステップにおける，μ_1^k と ϵ_n^k の更新値はそれぞれ式 (8)，(9) となる．w_n^k を式 (6)，σ_n^k を式 (7) で更新する．

$$\bar{\mu}_1^k = \frac{\sum_{n=1}^{N_k} \frac{h_n^k}{{\sigma_n^k}^2} \int_{-\infty}^{\infty} p(n,k|\omega,\theta) f(\omega)(\omega - \epsilon_n^k)\, d\omega}{\sum_{n=1}^{N_k} \frac{{h_n^k}^2}{{\sigma_n^k}^2} \int_{-\infty}^{\infty} p(n,k|\omega,\theta) f(\omega)\, d\omega} \tag{8}$$

$$\bar{\epsilon}_n^k = \frac{\nu^2 \int_{-\infty}^{\infty} p(n,k|\omega,\theta) f(\omega)(\omega - \mu_n^k)\, d\omega}{{\sigma_n^k}^2 + \nu^2 \int_{-\infty}^{\infty} p(n,k|\omega,\theta) f(\omega)\, d\omega} \tag{9}$$

---

*“Spectral Detection of Inharmonic Sound Sources in Mixed Sound based on Extended Harmonic Clustering using Maximum A Posteriori Estimation” by Seiya ODA, Hirokazu KAMEOKA, Takuya NISHIMOTO, and Shigeki SAGAYAMA (The University of Tokyo).

[1]式 (3) の偏微分を 0 と置くことで求まる．

3.2 比例係数パラメータ h_n^k の推定

3.1 節では楽器の種類が事前に分かっている場合を想定したが，楽器の種類あるいは h_n^k の理論値や経験的な値も不明な場合も考えうる．そこで，高調波周波数の基本周波数に対する比例係数 h_n^k 自体を事前分布によって確率的に制約を受けるモデルパラメータとして導入する．音 k の n 次高調波周波数 μ_n^k を，$h_n^k\mu_1^k$ で表す．また，事前分布は，さまざまな楽器音の情報から学習させて与えることが望ましいが，ここでは簡単のため平均 n，分散 ν の正規分布とする．以上より，MAP 推定により h_n^k を求めることができる．M ステップにおいて，MAP 推定により h_n^k を次式 (11) で更新し，最尤推定により μ_1^k を次式 (10)，w_n^k を式 (6)，σ_n^k を式 (7) で更新する．

$$\bar{\mu}_1^k = \frac{\sum_{n=1}^{N_k} \frac{h_n^k}{\sigma_n^{k\,2}} \int_{-\infty}^{\infty} p(n,k|\omega,\theta) f(\omega)\omega \; d\omega}{\sum_{n=1}^{N_k} \frac{h_n^{k\,2}}{\sigma_n^{k\,2}} \int_{-\infty}^{\infty} p(n,k|\omega,\theta) f(\omega) \; d\omega} \quad (10)$$

$$\bar{h}_n^k = \frac{\sigma_n^{k\,2} + \nu^2 n \mu_1^k \int_{-\infty}^{\infty} p(n,k|\omega,\theta) f(\omega)\omega \; d\omega}{\sigma_n^{k\,2} + \nu^2 n^2 \mu_1^{k\,2} \int_{-\infty}^{\infty} p(n,k|\omega,\theta) f(\omega) \; d\omega} \quad (11)$$

ただし，この定式化では，著しい非調和性楽器 (ドラムやトライアングルなど) は扱えないが，ピアノやティンパニなどのようにある程度ピッチ感を有するものに関しては十分有効であることが期待できる．

4 評価実験

提案手法の有効性を確認するため，(1) 調和性と仮定した場合のモデル，(2)h_n^k を理論値として固定したモデル，(3)h_n^k の理論値からの誤差をパラメータとして導入したモデル，(4)h_n^k 自体をパラメータとするモデルの間で比較検証を行った．

RWC 研究用音楽データベースに収録されている，非調和性楽器 (ピアノまたはトライアングル) とバイオリンの楽器音信号 (サンプリング周波数 44.1kHz) を人工的に加算し，多重音信号データを作成した．フレーム長 46ms の正規分布窓をかけて，周期 5ms で FFT を行い，短時間スペクトル系列を得た．開始フレームにおける各楽器音の μ_1^k の初期値は目視で与え，それ以降のフレームは直前フレームでの推定値を初期値することで基本周波数追跡を行った．また，h_n^k の理論値は文献 [4] を参照した．

また，検出した高調波周波数がどの程度正確であったかを定量的に評価するために，多重音信号のスペクトル系列から各楽器音ごとに推定したモデルの尤度分布と，各楽器音信号 (加算する前) のスペクトル系列を正規化したものとの間の KL(Kullback-Leibler) 情報量を基準とした．KL 情報量が 0 に近いほど，より正確にモデルを推定できたことを意味する．(3) では，ϵ_n^k の推定は非調和性楽器についてのみ行い，(4) では，h_n^k の推定は非調和性楽器，バイオリン両方について行った．

ピアノ音及びトライアングル音に対する (1)〜(4) それぞれに関する各時間の KL 情報量を図 1 上段，下段に示す．両者の結果より，(2) と比較して (3) でのモデル推定精度の向上が見られたので，誤差パラメータの導入がこれに寄与したと考えられる．トライアングルの場合は，(4) では必ずしも良い結果は得られなかったが，これはトライアングル音の非調和性が著しく，h_n^k を n を中心とした事前分布では原理的に扱える範囲ではないためであると考えられる．これに対し，ピアノ音の場合，(4) は (2) と (3) と近い結果が得られ，非調和性の度合がそれほど大きくない楽器音に対しては，(4) は十分有効であることが確認できた．

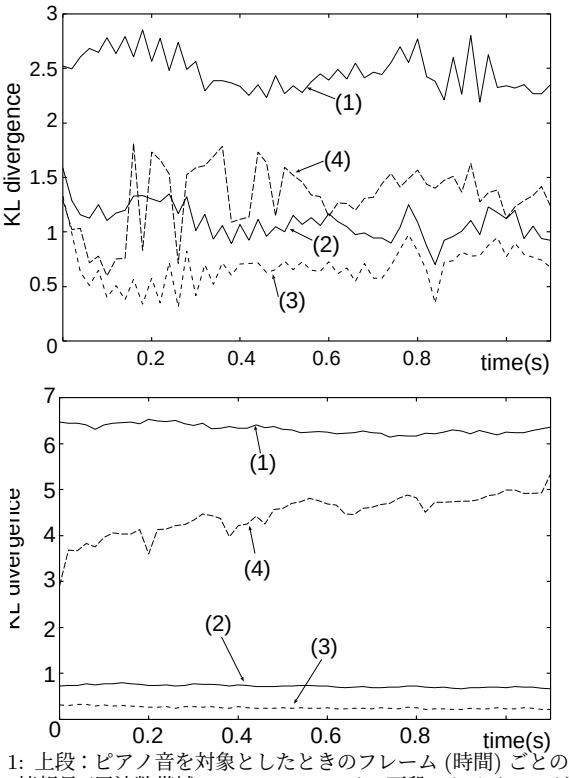

図 1: 上段：ピアノ音を対象としたときのフレーム (時間) ごとの KL 情報量 (周波数帯域 3.2kHz〜6.5kHz)．下段：トライアングル音を対象としたときのフレーム (時間) ごとの KL 情報量 (周波数帯域 0kHz〜6.5kHz)． (1) 調和性を仮定した場合のモデル (2)h_n^k を理論値として固定したモデル (3)h_n^k の理論値からの誤差をパラメータとして導入したモデル (4)h_n^k 自体をパラメータとするモデル

5 まとめ

本報告では，Harmonic Clustering を応用することで，非調和性楽器音の高調波周波数を高精度に検出する手法を提案した．実験により，提案手法の効果を確認した．今後は，モデルパラメータ ϵ_n^k 及び h_n^k の事前分布を学習して与え，精度向上を図る予定である．また，音響物理学などによって得られる知見をモデルに導入していきたい．

参考文献

[1] 柏野邦夫, 木下智義, 中臺一博, 田中英彦, “音楽情景分析の処理モデル OPTIMA における和音の認識,” 電子情報通信学会論文誌, Vol. J79-D-II, No. 11, pp. 1762–1770, 1996.

[2] M. Goto, “A Predominent-F0 Estimation Method for CD Recordings: MAP Estimation Using EM Algorithm for Adaptive Tone Models,” *Proceedings of the 2001 IEEE International Conference on Acoustics*, Speech and Signal Processing, ICASSP 2001, pp. V-3365–3368, 2001.

[3] 亀岡弘和, 西本卓也, 嵯峨山茂樹, “ハーモニック・クラスタリングによる多重音信号音高抽出における音源数とオクターブ位置推定,” 情報処理学会研究報告, 2003-MUS-51, pp.29-34, 2003.

[4] N.H. フレッチャー, T.D. ロッシング (岸憲史, 久保田秀美, 吉川茂訳), “楽器の物理学,” シュプリンガー・フェアラーク東京, 2002.