4S-7

倍音コーパスを用いた初期値依存性の低い多重基本周波数推定法

阪上 大地　　糸山 克寿　　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

1. はじめに

音響信号から楽譜を推定する自動採譜というタスクは，計算機を使用して音楽を理解する音楽情景分析における重要な要素技術である．その中心的な課題は，各時刻で鳴っている音の音程を推定する，多重基本周波数推定である．特に，ベイズ推論を用いた手法は，観測信号中に隠された楽器の種類，発音時刻，和音情報などの音楽的特徴量をモデル化できるため，音源分離 [1] や和音認識 [2] 等に応用しやすい．

潜在的調波配分法 (latent harmonic allocation; LHA) [3] はこのような手法の代表例であり，変分ベイズ法を用いた高速な推論が可能である．LHA では，観測音のスペクトログラムを単音のスペクトルの重ね合わせとして表現する．しかし，モデル中の各単音が任意の倍音構造を取りうるため，F0 や倍音振幅比として正解に近いものを与えなければ，不適切な倍音構造が推測され，誤った F0 が得られる．従来法では，この問題に対処するため事前分布を最適化する [3, 4] が，このアプローチは実在の楽器音に関する統計的根拠に基づいていないため，これらの手法の限界を予測し，応用や改善につなげることは困難である．したがって，楽器音に関する基本的な知識をもとに，初期値に対し頑健な手法を開発する必要がある．

本稿では，LHA に基づく多重基本周波数推定を初期値によらず頑健に行うため，MIDI 音源により作成した倍音コーパスを用いる手法を報告する．単音のスペクトルが高々第 M 次までの倍音を持つとすると，各倍音へのエネルギーの分配は $(M-1)$-シンプレックス上の座標として表現することができる．我々は楽器音の倍音構造として適切な領域がシンプレックス上の凸包領域であると仮定し，この凸包を倍音コーパスを用いて推定した．さらに，モデル中の全ての倍音構造がこの凸包に含まれるような推論手法を構成した．本手法の実験を様々な初期条件の元で行った結果，各音の倍音構造が適切かつ安定に推論可能であることを確認した．

2. 観測音のモデル化

D を観測スペクトログラムの時間フレーム数，F を周波数ビンの数とする．本稿では，LHA と同様に観測スペクトログラム X_{df} が時間フレームごとのヒストグラムの集合であると解釈する．すなわち，d 番目の時間フレームにおいて，f 番目の周波数 x_f が X_{df} 回観測されたと解釈する．

各時間フレームの観測スペクトルの生成モデルをあたえる．楽曲全体が高々 K 種類の楽器音の組み合わせで構成されていると考え，d 番目のヒストグラムを生成する確率分布 p_d が，K 個の楽器音の確率分布の重み付き和で表されると考える．

$$p_d(x|\pi_d,\theta)=\sum_{k=1}^{K}\pi_{dk}p_k(x|\theta_k) \tag{1}$$

Robust multiple F0 estimation against initialization using overtone corpus: *Daichi Sakaue, Katsutoshi Itoyama, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)*

ここで，π_{dk} は d 番目の時間フレームにおける各楽器音の割合である．p_k は k 番目の楽器音の確率分布，θ_k は k 番目の楽器音のモデルパラメータであり，これは次に述べるように混合ガウス分布などで表現できる．

2.1 楽器音のモデル化

ピアノやギターなどの楽器音を統計的手法によりモデル化する際には，楽器音の調波構造，すなわち基本周波数と倍音振幅比をモデルに適切に反映する必要がある．基本周波数は人間が知覚する楽器音の音程に対応し，倍音強度比は楽器音の音色と強い関係を持つ．LHA では，振幅スペクトル上で観測されるエネルギーの分布を対数周波数軸上の混合ガウス分布としてモデル化する．

$$p_k(x|\tau_k,\mu_k,\lambda_k)=\sum_{m=1}^{M}\tau_{km}\mathcal{N}(x|\mu_k+o_m,\lambda_k^{-1}) \tag{2}$$

ここで，μ_k は k 番目の音の基本周波数，λ_k はガウス分布の精度，τ_{km} は第 m 倍音の相対強度である．また，$\sum_m \tau_{km}=1$ である．$o_m=1200\log_2 m$ は第 m 倍音の相対位置である．

LHA に基づき多重基本周波数推定を精度よく行うためには，倍音振幅比を適切に推定する必要がある．例えば，$\mu_k=440$ Hz で，第 2, 4, 6 倍音のみ存在するという推定を行った場合，基本周波数は真値 (880 Hz) から 1 オクターブずれて推論されてしまう．

従来法では，この問題に対処するため，倍音構造に対する事前分布を与えている．しかし，このアプローチには次のような問題が存在する．適切な事前分布を手作業で与える手法 [4] の場合，最適なパラメータが観測音に含まれる楽器や音程の割合に依存するため，最適化が困難となる．また，階層的ベイズモデルを用いて事前分布そのものを推定する手法 [3] の場合，推定される結果は変分下限を最大化する事前分布であり，実在の楽器音に関する統計的根拠に基づいていない．

2.2 倍音コーパスを用いた定式化

MIDI 音源を用いて J 種類の楽器音を用意し，これらの楽器音の倍音構造（倍音テンプレート）を得る．τ^0_{jm} を j 番目の倍音テンプレートの第 m 倍音の割合とする．本手法では，楽器音は倍音テンプレートの重み付き和で表現可能な倍音構造のみをもつと仮定し，この重み（割合）を η_{kj} で表す．k 番目の音の確率密度関数は

$$p_k(x|\eta_k,\mu_k,\lambda_k)=\sum_{j,m=1}^{J,M}\eta_{kj}\tau^0_{jm}\mathcal{N}(x|\mu_k+o_m,\lambda_k^{-1}) \tag{3}$$

と書ける．このとき，k 番目の音の倍音振幅比 $\sum_j \eta_{kj}\tau^0_{jm}$ は，J 個の倍音テンプレート（倍音コーパス）の倍音振幅比によって張られる凸包に含まれる．

変分推論を行うため，観測 x_{dn} に対応する潜在変数 z_{dnkjm} を導入し，π, η, μ, λ に対し共役事前分布を与える．提案法の完全な同時分布は次のようになる．

$$p(X|Z,\mu,\lambda)=\prod_{dnkjm}\mathcal{N}(x_{dn}|\mu_k+o_m,\lambda_k^{-1})^{z_{dnkjm}} \tag{4}$$

$$p(Z|\pi,\eta) = \prod_{dnkjm} \left(\pi_{dk}\eta_{kj}\tau^0_{jm}\right)^{z_{dnkjm}} \quad (5)$$

$$p(\pi) = \prod_{d=1}^{D} \mathrm{Dir}(\pi_d|\alpha_0) \quad p(\eta) = \prod_{k=1}^{K} \mathrm{Dir}(\eta_k|\beta_0) \quad (6)$$

$$p(\mu,\lambda) = \prod_{k=1}^{K} \mathcal{N}(\mu_k|m_0,(\gamma_0\lambda_k)^{-1})\mathcal{W}(\lambda_k|w_0,\delta_0) \quad (7)$$

ここで，$\mathcal{N}$ は正規分布，$\mathcal{W}$ はウィシャート分布，Dir はディリクレ分布である．$\alpha_0,\beta_0,\gamma_0,\delta_0,m_0,w_0$ はハイパーパラメータである．推論には変分ベイズ法を用いる．

2.3 テンプレート数削減による高速化

本手法の計算時間は倍音テンプレート数 J に比例するため，不要なテンプレートを削減することで高速化できる．凸包の内点の座標は頂点の座標の非負線形和として書けるため，内点にあたるテンプレートは冗長であり，削除しても凸包は変化しない．この操作を近似的に行うため，

$$\tilde{J} = \cup_m \left\{\mathrm{argmin}_j \tau^0_{jm}, \mathrm{argmax}_j \tau^0_{jm}\right\} \quad (8)$$

とする．推論精度を上げるため，二番目に大きなテンプレートと二番目に小さなテンプレートも $\tilde{J}$ に加える．

3. 実験及び考察

LHA と提案法の初期値に対する頑健性を調べるため，20 曲の楽曲と 3 種類の初期化法のもとで多重基本周波数推論を行い，結果を比較した．

3.1 コーパスの作成

我々はまず，MIDI 音源 (Roland SD-80) を用いて倍音コーパスを作成した．このとき，MIDI のプログラム 1 番から 80 番までの楽器から適切な調波構造を持たない 10 楽器を除いた 70 楽器を用いた．これらの楽器音を 440 [Hz] で 1 秒間再生し，得られた信号をウェーブレット変換して倍音コーパスを作成した．さらに，2.3 節に書いた方法を用いて冗長な倍音テンプレートを取り除いた．

3.2 観測スペクトログラムの作成

実験には，RWC 音楽データベース [5] から，ピアノソロ 5 曲 (Jazz, No. 1–5,) ギターソロ 5 曲 (Jazz, No. 6–10,) 室内楽 10 曲 (Classic, No. 12–21) を使用した．各楽曲は，MIDI 音源 (YAMAHA MOTIF-XS) を用いて先頭 24 秒を録音し，ウェーブレット変換を行った．

3.3 実験条件

様々な初期値のもとでの推論の頑健性を調べるため，三種類の初期化法を用いて実験を行った．これらを順に，乱数初期化，線形初期化，指数初期化と呼ぶ．乱数初期化では，負担率を一様乱数で初期化した．これは一般に考えうる最悪のケースとみてよい．残る二つの初期化法では，基本周波数の初期値を C1 から C7 に設定し，分散は 50 [cent] とした．線形初期化では，τ_{km} または η_{kj} が一様となるように初期化し，指数初期化では倍音重みの初期値が 2^{-m} に比例する形で初期化した．実験の際，従来法と提案法の事前分布は全て無情報とした．

変分ベイズ法による推論の後，時間フレームごとに一定の閾値を超える N_{dk} を結果として出力した．手法の有効性を判断するため，MIDI ファイルから各時間フレームの真の基本周波数の組を計算し，推定結果との時間フレームごとの F 値に基づき判断した．また，両手法の潜在的な最高性能を比較するため，各手法や楽曲等に対し閾値を個別に最適化した．

表 1: 基本周波数の F 値．太字は最大値を表す．

	従来法 (LHA)			提案法		
Genre	乱数	線形	指数	乱数	線形	指数
Piano	31.1	51.3	58.5	55.0	**62.3**	58.1
Guitar	12.8	48.8	**76.7**	64.3	73.9	72.6
Chamber	23.8	36.1	49.1	46.8	**53.6**	50.9

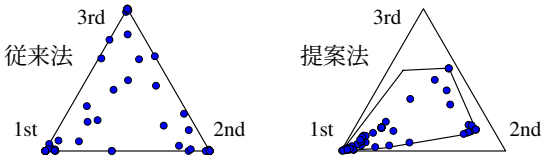

図 1: RM-C012 に対し，乱数初期化で推定された第 3 倍音までの相対振幅比を示す．右図中に推論に使用した凸包を示す．対応する F 値は 19.4 (LHA) および 63.4 (提案法) である．

3.4 実験結果

実験結果を表 1 に示す．LHA に指数初期化を用いた場合と提案法に線形初期化を用いた場合の性能が似ていることから，適切な初期値を与えた場合 LHA の推論が上手く行えることがわかる．一方，LHA に乱数初期化や線形初期化を用いた場合，指数初期化に比べて性能が大幅に悪化する．これは，LHA の倍音構造のモデル化が万能でなく，初期値依存性が存在することを示している．提案法は三種類の初期化条件全てについて安定した性能を達成しているため，初期値に対して頑健である．推論結果の一例を図 1 に示す．

3.5 考察

本稿では，楽器音として適切な倍音構造の範囲が凸包であると仮定して議論を進めた．実際，ある音程を持つ楽器音の倍音構造を集めて凸包を作ると，凸包の内点に対応する倍音構造は凸包の頂点に対応する倍音構造の線形和として得られる．これは，同一のピッチの音が同時に複数鳴った時，聴感上得られるピッチが元のピッチと同じであることに対応している．

4. おわりに

本稿では，MIDI 音源を用いて適切な倍音構造の範囲を推定し，LHA の観測モデルの倍音振幅比をこの領域内に閉じ込める初期値に対し頑健な多重基本周波数推論法を報告した．今後は，倍音構造等をさらに精密にモデル化し，性能の高い推論アルゴリズムを開発していきたい．なお，本研究は科研費 (S), GCOE の支援を受けた．

参考文献

[1] K. Itoyama *et al.*: Parameter estimation for harmonic and inharmonic models by using timbre feature distributions, 情処論, vol.50, no.7 (2009), 1757–1767.

[2] Y. Ueda *et al.*: HMM-based approach for automatic chord detection using refined acoustic features, *ICASSP 2010*, pp. 5518–5521, 2010.

[3] 吉井 他: 多重音基本周波数解析のための無限潜在的調波配分法, 情処研報, 2010-MUS-86, 2010.

[4] H. Kameoka *et al.*: A multipitch analyzer based on harmonic temporal structured clustering, *IEEE Trans. on ASLP*, vol.15, no.3, pp. 982–994, 2007.

[5] 後藤 他: RWC 研究用音楽データベース，情処論, vol.45, no.3 (2004), 728–738.