ハーモニッククラスタリングによる多重音の基本周波数推定*

◎亀岡弘和　西本卓也　篠田浩一　嵯峨山茂樹　（東京大学大学院情報理工学系研究科）

1　はじめに

音声や楽器音などの複数音源または単一音源による複数の音響信号が混在したものを多重音という．音楽演奏における多重音の基本周波数推定手法はこれまでにスペクトルのテンプレートマッチングを用いたもの [1, 2]，櫛形フィルタを用いたもの [3]，混合正規分布を用いたもの [4] などが提案されてきた．従来の多重音の基本周波数推定手法では，同時発音数の増加や音域の拡大に伴う計算量の組み合わせ的爆発，自然楽器の音色変動の影響，ビブラートなどのようなピッチ変動の影響などの問題 [1, 2, 3] があった．我々はこれらの問題を解決する基本周波数推定手法を目指し，新しいアルゴリズム“ハーモニッククラスタリング”を提案する．本報告ではまずこのアルゴリズムの原理について述べる．また，今回は MIDI 音源による音楽信号を対象とし，同時発音数は既知として，本手法を用いた基本周波数推定の性能の検証を行った．

2　ハーモニッククラスタリング

単一音は，一般に基本周波数成分とその整数倍の周波数の倍音成分をもつ．(ベルやティンパニのように非調和性が著しい楽器もあるが，ここでは扱わない．) 多重音のスペクトルは倍音構造をもつスペクトルの重ね合わせとなる．そこで，基本周波数成分とその倍音成分とをまとめて 1 つのクラスタとして扱うようなクラスタリング (ハーモニッククラスタリング) を行うことで複数の基本周波数を推定する手法を検討した．以下に，定式化を示す．

2.1　k-means アルゴリズムの適用

まず，直観的な説明のために k-means アルゴリズムに基づく方法を述べる．離散的周波数 ω_i $(i = 1, \cdots, I)$ について信号のパワースペクトル密度 $f(\omega_i)$ が与えられているとする．また，同時発音数 K(クラスタ数) は既知とする．

ステップ 0:　各基本周波数に対応するセントロイド $\boldsymbol{\mu} = \mu_1, \cdots, \mu_k, \cdots, \mu_K$ の初期値を与える．

ステップ 1:　μ_k $(k = 1, \cdots, K)$ とそれらの整数倍の値 $n\mu_k$ $(n = 1, 2, 3, \cdots)$ を重心周波数と呼び，スペクトルの周波数軸上にとる．離散的周波数 ω_i それぞれについてユークリッド距離が最も小さい重心周波数 $n\mu_k$ を探し，ω_i をクラスタ k の n 倍音成分として分類する．これは図 1 のように，すべての重心周波数 $n\mu_k$ $(n = 1, 2, \cdots; k = 1, \cdots, K)$ について周波数軸上で隣接する他の重心周波数との中点で領域を分割し，領域を T_n^k とすることに相当する．

ステップ 2:　以下のような目的関数を最小化するように μ_k を $\hat{\mu}_k$ に更新する．

$$D(\boldsymbol{\mu}) = \sum_{k=1}^{K} \sum_{n} \sum_{\omega_i \in T_n^k} d(\omega_i, n\mu_k) f(\omega_i) \qquad (1)$$

ただし，f_{ω_i} は ω_i に対応するパワースペクトル成分を表し，$d(\omega_i, n\mu_k)$ は周波数 ω_i と周波数 $n\mu_k$ とのユークリッド距離の二乗である．$\hat{\mu}_k$ は以下により得られる．

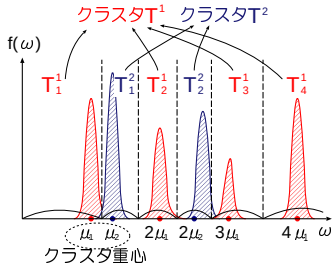

図 1: ハーモニッククラスタリング

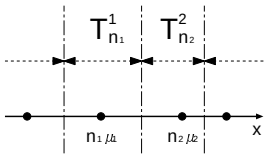

図 2: 矩形領域

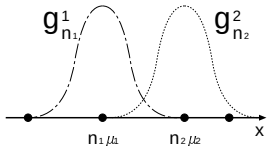

図 3: 領域の確率的分割

$$\hat{\mu_k} = \sum_{n} n \sum_{\omega_i \in T_n^k} \omega_i f(\omega_i) \Big/ \sum_{n} n^2 \sum_{\omega_i \in T_n^k} f(\omega_i) \qquad (2)$$

以上のステップ 1 と 2 の反復計算による μ_k の収束値が本手法における基本周波数推定値である．μ_k の収束は保証されているが，誤った局所解に陥る可能性があるため，初期値によってはこの値が必ずしも実際の基本周波数と等しくなるとは限らない．

2.2　アルゴリズムの拡張

前節で述べたハーモニッククラスタリングに，次のような拡張を行う．

1. スペクトルを離散分布 (度数分布) ではなく，連続分布 (度数密度分布) として考える．また，周波数も連続的に扱う．$(\omega_i \to \omega)$
2. 線形周波数軸ではなく対数周波数軸を用いる．
3. T_n^k を矩形領域 (図 2) により排他的に分割するのではなく，共有を許して確率的に分割する (図 3)．
4. 距離尺度 $d(\omega, n\mu_k)$ として対数正規尤度を用いる．

$w_n^k(\omega)$ をすべての重心周波数 $n\mu_k$ を平均とする正規分布の混合分布の重みとすると，1.〜4. より式 (1) は $w_n^k(\omega)$ を用いて以下のように書き直せる．ただし，$f(\omega)$ はパワースペクトル密度を表す．

$$D(\boldsymbol{\mu}) = \sum_{k=1}^{K} \sum_{n} \int_{-\infty}^{\infty} d(\omega, n\mu_k) w_n^k(\omega) f(\omega) d\omega \qquad (3)$$

$$d(\omega, n\mu_k) = \frac{1}{2} \left\{ \log(2\pi) + \log\sigma^2 + \frac{(\log\omega - \log(n\mu_k))^2}{2\sigma^2} \right\}$$

式 (3) の最小化問題は混合正規分布における EM アルゴリズムと同様の定式化となるので，w_n^k は，次式となる．

$$w_n^k(\omega) = \frac{g_n^k(\omega)}{\sum_k \sum_n g_n^k(\omega)} \qquad (4)$$

$$g_n^k(\omega) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(\log\omega - \log(n\mu_k))^2}{2\sigma^2} \right\} \qquad (5)$$

* “Multipitch Estimation using Harmonic Clustering” by Hirokazu KAMEOKA, Takuya NISHIMOTO, Koichi SHINODA and Shigeki SAGAYAMA (Graduate School of Information Science and Technology, The University of Tokyo).

表 1: 初期セントロイドの収束範囲の評価

対象音		初期値		半オクターブ	
音名	基本周波数	下限	上限	下	上
F4	740	668	980	523	1047
E4	659	549	980	466	932
D4	587	506	969	415	830
C#4	554	463	958	392	783
B3	494	420	775	349	699
A3	440	366	689	311	622

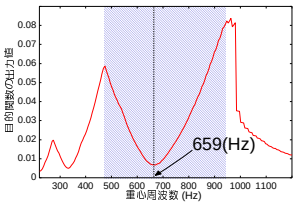

図 4: 目的関数値

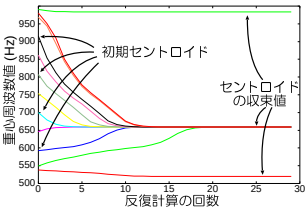

図 5: セントロイドの収束

また，$\log \mu_k$ は次式のような $\log \hat{\mu_k}$ により更新する．

$$\log \hat{\mu_k} = \frac{\sum_n \int_{-\infty}^{\infty} (\log \omega - \log n) w_n^k(\omega) f(\omega) d\omega}{\sum_n \int_{-\infty}^{\infty} w_n^k(\omega) f(\omega) d\omega} \quad (6)$$

本手法の利点として次のようなことが挙げられる．クラスタ数が，推定可能な最大同時発音数となるため，同時発音数が多い対象音に対しても組み合わせ的爆発は起こらず，たかだかクラスタ数に比例した計算量で済む．推定に倍音比の情報を用いないため，音色の不規則な時間変動の影響を受けない．また，ビブラートなどのようなピッチ変動にも追従できる．ただし，適切な初期セントロイドの設定が必要である．

3 実験評価

本実験では，ハーモニッククラスタリングにより正しい推定値に収束するための初期セントロイドの収束範囲の評価と，正しい基本周波数値が開始フレームに与えられたときに以後のピッチトラッキングの精度の評価を行った．

3.1 初期セントロイドの収束範囲

MIDI 音源 (ヴァイオリン音) を対象として用いた．フレーム長 100(ms)，ハミング窓を用いた FFT により周波数解析を行なった．

6 種類の単音 (A3,B3,C#4,D4,E4,F4) それぞれについて，正しい値に収束する初期セントロイドを 1.08[Hz] おきに調べ，収束範囲の上限と下限の各単音につき 10 フレーム分の平均をとったものを表 1 に示す．また，参考のために各単音から半オクターブ下と半オクターブ上の周波数を示す．数値の単位はすべて [Hz] である．

図 4 は，単音 (E4，基本周波数：659[Hz]) のスペクトルを用いたときの目的関数 $D(\boldsymbol{\mu})$ の異なる μ_k についての値を示す．斜線部の周波数の範囲が初期セントロイドの収束範囲となる．また図 5 は，それぞれの初期セントロイドについて反復計算ごとの更新値を表す．

3.2 ピッチトラッキング

基本周波数の推定値を次のフレームにおける処理の初期セントロイドとすることにより，ピッチの連続的な変化を逐次的に追跡することができる．3.1 と同様に，フレーム長 100(ms)，ハミング窓を用いた FFT により周波数解析を行ない，フレームシフトは 10(ms) とした．提案手法を用い，MIDI 音源 (ヴァイオリン音) による音楽信号を対象として用いて同時発音数を既知としたピッチトラッキング性能の検証を行った．開始フレームにおける基本周波数値は既知とし，これを開始フレームにおける初期セントロイドとして与えた．図 6 に同時発音数が 1 の単旋律，図 7 に同時発音数 2 の複旋律の基本周波数の追跡結果の一例を示す．音楽信号からの基本周波数推定の正解率は，単旋律 94(%)，複旋律 (同時発音数は 2)88(%) であった．ピッチトラッキングの誤推定は，3.1 で得られた収束範囲外にピッチが跳躍するときに見られた．

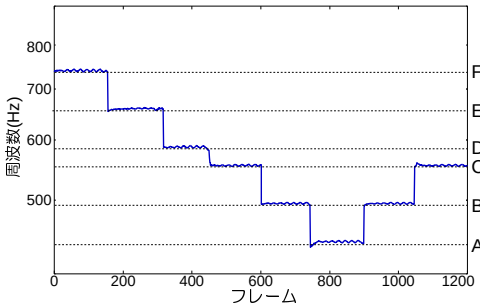

図 6: 単旋律のピッチトラッキング結果例

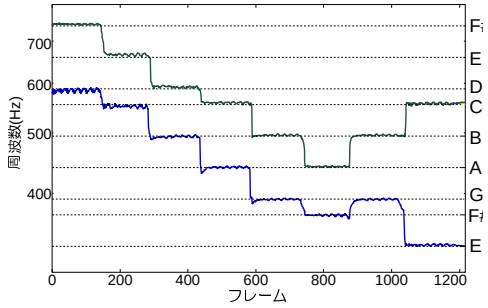

図 7: 複旋律のピッチトラッキング結果例

4 おわりに

本報告では，反復計算による多重音の基本周波数の推定手法 “ハーモニッククラスタリング” を提案した．ハーモニッククラスタリングにより正しい推定値を得ることができる初期セントロイドの収束範囲を評価を行った．また，MIDI 音源による音楽信号のピッチトラッキング性能の検証を同時発音数を既知として行った．実験において，逐次的に基本周波数が追跡できた．

今後は，適切な初期セントロイドの決定方法や，同時発音数の推定方法の検討を行い，さまざまな条件で動作を確認するとともに実音楽信号 (多重音) に適用したい．

謝辞

本研究の一部は，科学技術振興事業団戦略的基礎研究推進事業 (CREST) (「脳を創る」聴覚脳研究プロジェクト) の支援を受けて行われた．

参考文献

[1] 中臺一博, 柏野邦夫, 田中英彦: “音楽音響信号を対象とする音源分離システム,” 情報処理学会技術研究報告, SIGMUS1-1, pp. 1–8, 1993.
[2] 小野徹太郎, 斎藤英雄, 小沢慎治: “自動採譜のための GA を用いた混合音推定,” 計測自動制御学会論文集, Vol. 33, No. 5,
[3] 三輪多恵子, 田所嘉昭, 斎藤努: “くし形フィルタを利用した採譜のための異楽器音中のピッチ推定,” 電子情報通信学会論文誌, Vol. J81-D-II, No. 9, pp. 1965–1974, 1998.
[4] 後藤真孝: “音楽音響信号を対象としたメロディーとベースの音高推定,” 情報処理学会研究報告, 99-MUS-31-16, Vol. 99, No. 68, 1999.