

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アルゴリズム

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

〒113-0033 東京都文京区本郷7-3-1

^{††} 東京工業大学情報理工学系研究科

〒152-8552 東京都目黒区大岡山2-12-1

E-mail: †{kameoka,nishi,sagayama}@hil.t.u-tokyo.ac.jp, ††shinoda@cs.titech.ac.jp

あらまし 本報告では、複数音源による多重音のモノラル音響信号に対して、基本周波数を推定する手法を提案する。1つのクラスタが、基本周波数の近傍および倍音周波数の近傍の周波数帯域によって構成されると見なすことにより、クラスタ内のスペクトル成分のクラスタリングを行うことができる。このようなアプローチを、ハーモニック・クラスタリングと呼ぶ。個々のスペクトル成分をいずれか1つのクラスタのみに割り当てるのではなく、複数のクラスタが確率的な尺度により共有するように扱うことで、このクラスタリングはEMアルゴリズムとして定式化される。本手法を用い、独唱および多声音楽のモノラル音響信号から基本周波数推定を行い、性能の検証を行った。倍音構造の変化が著しい無伴奏歌唱に対し90.6%の音名正解率を得た。また、ビブラートやグリッサンドのようなピッチ変動が含まれるヴァイオリン三重奏に対し83.2%の音名正解率を得た。

キーワード 多重音ピッチ推定, k -means アルゴリズム, 混合正規分布, EM アルゴリズム

Multi-Pitch Estimation Using Harmonic Clustering

Hirokazu KAMEOKA[†], Takuya NISHIMOTO[†], Koichi SHINODA^{††}, and Shigeki SAGAYAMA[†]

[†] Graduate School of Information Science and Technology, University of Tokyo

7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033 Japan

^{††} Graduate School of Information Science and Engineering, Tokyo Institute of Technology

2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8552 Japan

E-mail: †{kameoka,nishi,sagayama}@hil.t.u-tokyo.ac.jp, ††shinoda@cs.titech.ac.jp

Abstract This report deals with a method for estimating the fundamental frequency of monaural audio signals containing sounds of various sound sources. As considering that a single cluster consists of the ranges close to the fundamental frequency and the respective harmonic frequencies, the clustering can be applied according to the spectral components within the cluster. We call this approach "Harmonic Clustering". It is further formulated as EM algorithm not by classifying each frequency to a single cluster but by assuming that multiple clusters can probabilistically share the same frequency. We experimented on estimating the fundamental frequencies of monaural musical audio signals using the proposed method. The accuracy of the pitch estimation for an unaccompanied song, which generally has drastic spectral changes, was 90.6% and for a violin trio, which often has pitch fluctuations mainly caused by vibratos or glissandi, was 83.2%.

Key words multi-pitch estimation, k -means algorithm, gaussian mixture distribution, EM algorithm

1. ま え が き

音声や楽器音などの複数音源の音響信号が混在したものを多重音という。多重音の基本周波数推定の研究は、音楽情報科学の分野では自動採譜や音源分離のための重要な要素の一つとし

て取り上げられている。

自動採譜や音源分離の研究はこれまでに数多く試みられており、1990年頃より複数音源による多重音または多声音楽を対象とされてきた[1]~[5]。多重音のスペクトルは、音源同士の基本周波数成分や倍音成分が互いに複雑に重なり合うため、観測

されたスペクトルから多重音に含まれる音源数や各音源の基本周波数は単純には推定できない。

この問題に対し、従来より楕円フィルタを用いたアプローチが提案されている [6]。楕円フィルタを用いた手法では、周波数領域において等間隔に零点を有するフィルタの次数を変化させて非零出力を検出することで多重音の音名を順次判別できる。この手法は、演奏楽器の種類や楽器音の倍音比に依存せず基本周波数の検出が可能である。また、近年では、単一音の倍音構造を確率分布としてモデル化し、その確率分布モデルの混合分布の重みを推定する手法が提案されている [7]~[9]。これは、あらゆる基本周波数の単一音が同時発音をしていると見なして相対的に優勢な基本周波数を推定する手法であり、高い分解能で基本周波数を得ることができる。我々は、前者のように倍音比のモデリングを必要とせず、さらに後者のように高い分解能で基本周波数の推定が可能な手法、“ハーモニック・クラスタリング”を提案し、既に予備実験を行った [10]。本報告では、まずこの手法の原理とアルゴリズムについて述べ、これを用いて歌唱および多声音楽の音響信号を対象としたピッチトラッキングの実験を行い、性能の評価を行う。

2. 問題の定式化

2.1 周波数のクラスタリング

まず、本問題の定式化の概要を述べる。単一正弦波（純音）のスペクトルは本来は線スペクトルであるが、短時間分析では周波数に幅を持つスペクトルとして観測される。本報告では、これを周波数軸上のクラスタ（離散化した周波数点ごとにパワースペクトルに対応する頻度を持つクラスタ）と考えて、そのクラスタ重心を求めることによって、周波数を推定する。この考え方をまず単一音の倍音構造、次に多重音へと拡張し、 k -means アルゴリズムに基づくハーモニッククラスタリングと呼ぶ周波数推定アルゴリズムを導く。さらに、周波数軸を対数尺度に、周波数軸距離関数を正規分布の対数尤度に、クラスタ帰属判定を確率的に置き換えることによって、この問題を EM アルゴリズムの問題として定式化する。以下、順に説明する。

2.2 線スペクトルの広がり

スペクトル解析手法では、一般に短時間の解析区間に窓関数をかける。そのため、純音の信号であっても線スペクトルではなく、その周波数を中心に左右に広がりをもつスペクトルが観測される。これは、窓関数のフーリエ変換と線スペクトルとの畳み込みを行うことに相当するためである。一般に用いられる窓関数においては、ある周波数の信号のスペクトルがその周波数を中心として左右に単調減少、かつ左右対称な形状として観測される。また、解析区間内で周波数が連続的に変化する信号からは、その変化にまたがるような広がりをもつスペクトルが観測される。上記の解析手法により、スペクトルが広がりをもつ形状として観測されるのはこれらに起因する。そこで、任意の信号の短時間スペクトルを、次の仮定を満たす信号のスペクトルとして見立てる。

(仮定 1) 解析区間内では周波数 Ω は一定である。

(仮定 2) 解析区間内でエネルギーは一定である。

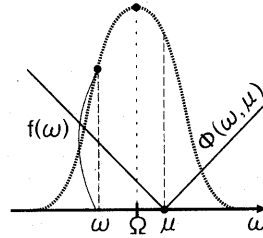


図 1 純音の Ω の推定

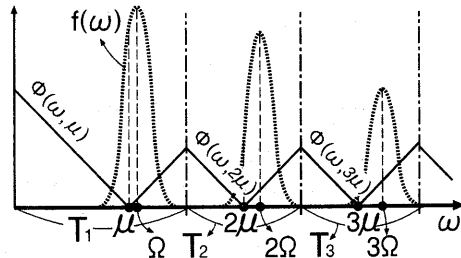


図 2 倍音構造をもつ単一音の Ω の推定

上の仮定は、短い解析区間において連続的に周波数が変化する場、その変化を、微小である、あるいは直線近似できると仮定することに相当する。(直線近似すれば、 Ω は区間内における周波数の平均と見なすことができる。) 以上より、スペクトルは Ω を中心として左右に単調減少、かつ左右対称な形状として観測される。観測スペクトルのピーク ($= \Omega$) は次のような距離関数

$$D_1(\mu) = \int_{-\pi}^{\pi} \varphi(\omega, \mu) \cdot f(\omega) d\omega \quad (1)$$

を最小化する μ を求めることで得られる (図 1)。ただし、 $-\pi \leq \omega \leq \pi$ を周波数、 $\varphi(\omega, \mu)$ を ω と μ との距離尺度を表す μ を中心とした左右対称な関数、 $f(\omega)$ を観測スペクトル密度とする。実際の計算では、 ω_i ($i = 1, 2, \dots, I$) を離散的周波数 (ω_i をナイキスト周波数とする) として、上記の式は積和で計算する。

2.3 倍音構造

一般に、単一音信号は 1 つの周波数成分だけではなく、基本周波数と倍音周波数の成分からなる。そこで、このような信号に関してさらに次のような仮定をする。

(仮定 3) 倍音周波数は基本周波数の整数倍である。

つまり、ベルやティンパニなどのように非調和性の著しい楽器音はここでは扱わない。(ただし、これは単に式を簡略化するための仮定であり、必ずしも必要ではない。) 以上の 3 つの仮定に従えば、単一楽器音の基本周波数 Ω は、式 (1) と同様な考え方で以下の距離関数 $D_2(\mu)$ を局所最小化する μ を求めることで得られる (図 2)。 n を自然数とする。

$$D_2(\mu) = \sum_n \int_{T_n} \varphi(\omega, n\mu) \cdot f(\omega) d\omega \quad (2)$$

ただし、 T_n は、以下のような集合とする。

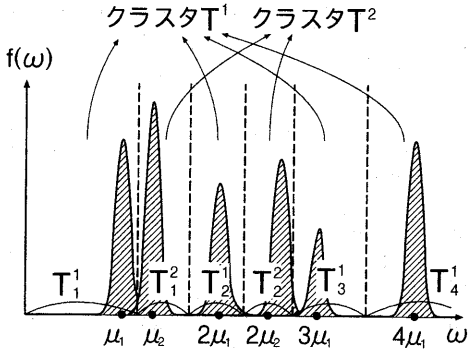


図3 クラスタ数が2つの場合のクラスタ帯域

$$T_n = \left\{ \omega \mid n = \operatorname{argmin}_m |\omega - m\mu| \right\} \quad (3)$$

すなわち、 T_n は図2のように $n\mu$ と $(n-1)\mu$ との中点および $n\mu$ と $(n+1)\mu$ との中点を両端とした帯域に相当する。

3. k-means アルゴリズムとしての定式化

以上のアプローチは、複数音源による多重音信号に対しても同様に適用できる。音源数を K 個とすると、 $\boldsymbol{\mu} = \{\mu_1, \dots, \mu_K\}$ を用いて目的関数 $D_3(\boldsymbol{\mu})$ を以下のように定義する。

$$D_3(\boldsymbol{\mu}) = \sum_k \sum_n \int_{T_n^k} \varphi(\omega, n\mu_k) \cdot f(\omega) d\omega \quad (4)$$

$$T_n^k = \left\{ \omega \mid (n, k) = \operatorname{argmin}_{(m, l)} |\omega - m\mu_l| \right\} \quad (5)$$

この目的関数の局所最小化問題は、 μ_k をクラスタ重心と見なすことで $T^k = \{T_1^k \cup T_2^k \cup \dots\}$ をクラスタ帯域 k として扱うことができ(図3)、k-means アルゴリズムにより定式化できる。

k-means アルゴリズムは、クラスタ境界の更新とクラスタ重心の更新の2つのステップを繰り返すことでサンプルデータのクラスタリングを行う手法である。サンプルデータを $\boldsymbol{x} = \{x_1, \dots, x_J\}$ 、クラスタ重心を $\boldsymbol{\mu} = \{\mu_1, \dots, \mu_K\}$ 、クラスタ帯域を T^k とすると、目的関数 $D_4(\boldsymbol{\mu})$ は以下となる。

$$D_4(\boldsymbol{\mu}) = \sum_k \sum_{x_j \in T^k} \varphi(x_j, \mu_k) \quad (6)$$

$$T^k = \left\{ x_j \mid k = \operatorname{argmin}_l |x_j - \mu_l| \right\} \quad (7)$$

式(4)は、 ω を値とするサンプルデータが $f(\omega)$ に比例した個数分だけ分布していると考えることにより、式(6)と同型になる。このとき、対象とする周波数帯域を離散化した $\{\omega_i\}$ を用いてサンプルデータ \boldsymbol{x} を概念的に表現すると、以下ようになる。

$$\boldsymbol{x} = \left\{ \underbrace{\omega_1, \dots, \omega_1}_{f(\omega_1) \text{ 個}}, \underbrace{\omega_2, \dots, \omega_2}_{f(\omega_2) \text{ 個}}, \dots, \underbrace{\omega_i, \dots, \omega_i}_{f(\omega_i) \text{ 個}}, \dots, \underbrace{\omega_J, \dots, \omega_J}_{f(\omega_J) \text{ 個}} \right\}$$

クラスタリングの原理に基づいて基本周波数を推定するこの提案手法を我々はハーモニック・クラスタリングと呼ぶ。

以下にハーモニック・クラスタリングの手順について述べる。

ステップ0: (初期設定)

クラスタ重心 $\boldsymbol{\mu} = \{\mu_1, \dots, \mu_K\}$ の初期値を与える。

ステップ1: (クラスタ帯域 T^k の更新)

μ_k ($k = 1, \dots, K$) とそれらの整数倍の値 $n\mu_k$ ($n = 1, 2, 3, \dots$) を重心周波数と呼び、スペクトルの周波数軸上にとる。それぞれの ω_i 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 倍音帯域 T_n^k として分類する。すなわち、すべての重心周波数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帯域を矩形分割し、 $n\mu_k$ が属する区間を T_n^k とすることに相当する。

ステップ2: (クラスタ重心 $\boldsymbol{\mu}$ の更新)

すべての k について μ_k を $\bar{\mu}_k$ に更新する。 $\varphi(\omega_i, n\mu_k) = (\omega_i - n\mu_k)^2$ とした場合、 $D_3(\boldsymbol{\mu})$ を最小化する $\bar{\mu}_k$ は次式となる。更新後、ステップ1に戻る。

$$\bar{\mu}_k = \frac{\sum_n \sum_{\omega_i \in T_n^k} \omega_i f(\omega_i)}{\sum_n \sum_{\omega_i \in T_n^k} f(\omega_i)} \quad (8)$$

式(5)より、ステップ1においても $D_3(\boldsymbol{\mu})$ を最小化するため、 $\boldsymbol{\mu}$ の収束性は保証される。そこで、 $\boldsymbol{\mu}$ の収束値を推定基本周波数とする。ただし、誤った局所解に陥る可能性があるため、正しい基本周波数に収束させるには適切な初期設定(ステップ0)を行う必要がある。

4. EM アルゴリズムとしての定式化

4.1 重複する倍音成分を確率的に分離するクラスタリング

以上のアルゴリズムではクラスタ帯域を矩形的に分割するため、音源同士の倍音成分が重複する場合においても、その周波数の帯域はいずれか1つのクラスタ帯域の一部と見なした。本章では、矩形的な帯域分割をするのではなく、周波数の帯域を複数のクラスタが共有できるように確率的な帯域分割を行い、重複する倍音成分を分離するクラスタリングについて述べる。

まず、 ω_i がクラスタ k の n 倍音帯域に属する確率を $p_n^k(\omega_i)$ とする。また、 ω_i と $n\mu_k$ の距離尺度を表す関数 $\varphi(\omega_i, n\mu_k)$ を、 $n\mu_k$ を平均とし、重み w (一定) をかけた正規分布の ω_i における対数尤度とする。また、基本周波数と倍音周波数の揺らぎの尺度を正規化するため、以後周波数軸を対数スケールで扱い、 $x_j = \log x_j$ 、 $\omega'_i = \log \omega_i$ 、 $\mu'_k = \log \mu_k$ と表現する。観測スペクトルを周波数方向に対数変換したものを $f'(\omega'_i)$ で表す。上述により、目的関数 $D_3(\boldsymbol{\mu})$ は以下ようになる。

$$D_5(\boldsymbol{\mu}') = \sum_k \sum_n \sum_i \varphi(\omega'_i, \mu'_k + \log n) \cdot p_n^k(\omega'_i) \cdot f'(\omega'_i)$$

$$\varphi(\omega'_i, \mu'_k + \log n) = \log \frac{w}{\sqrt{2\pi\sigma^2}} - \frac{(\omega'_i - \mu'_k - \log n)^2}{2\sigma^2} \quad (10)$$

4.2 倍音構造モデルの混合モデルの最尤推定

ここでは倍音構造の拘束をもった混合正規分布の μ' の最尤推定法について述べる。平均 $\{\mu'_k, \mu'_k + \log 2, \dots, \mu'_k + \log n, \dots\}$ 、分散 σ 、重み w の混合正規分布である倍音構造モデルを K 個混合した分布における、サンプルデータ $\mathbf{x}' = \{x'_1, \dots, x'_j\}$ に対する尤度は式 (11) となる。 w は倍音構造モデルの混合分布を正規化するための共通な係数である。

$$P(\theta) = \prod_j \left\{ \sum_k \sum_n w N(x'_j | \mu'_k + \log n, \sigma) \right\} \quad (11)$$

$$N(x | \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} \quad (12)$$

$$\sum_k \sum_n w = 1 \quad (13)$$

$$\theta = \{\mu', \sigma (\text{固定})\} \quad (14)$$

EM アルゴリズムにおける Q 関数は以下ようになる。

$$Q(\theta, \bar{\theta}) = \sum_j \sum_{k,n} P_{\bar{\theta}}(n, k | x'_j) \log w N(x'_j | \bar{\mu}'_k + \log n, \sigma) \quad (15)$$

$$= \frac{1}{F} \sum_i \sum_{k,n} P_{\bar{\theta}}(n, k | \omega'_i) f'(\omega'_i) \log w N(\omega'_i | \bar{\mu}'_k + \log n, \sigma) \quad (16)$$

$$P_{\bar{\theta}}(n, k | x'_j) = \frac{N(x'_j | \mu'_k + \log n, \sigma)}{\sum_k \sum_n N(x'_j | \mu'_k + \log n, \sigma)} \quad (17)$$

ただし、式 (16) の F はスペクトル密度を確率密度分布として扱うための正規化係数である。2章における (仮定 1)、(仮定 2) の代わりに次のような仮定をする。

(仮定 4) スペクトルの広がりには、正規分布で近似できる。

この仮定に従えば、 $p_n^k(\omega'_i) = P_{\bar{\theta}}(n, k | \omega'_i)$ となり、正規化係数 F はパラメータ推定に関して影響を及ぼさないので式 (9) と式 (16) は等価と考えることができる。従って、次の手順により混合正規分布のパラメータ推定を行う。

ステップ 0: (初期設定)

クラスタ重心 $\mu' = \{\mu'_1, \dots, \mu'_K\}$ の初期値を与える。

ステップ 1: (E ステップ)

式 (12) により $p_n^k(\omega'_i)$ を算出し、 $D_5(\mu') (= Q(\theta, \bar{\theta}))$ を求める。

ステップ 2: (M ステップ)

μ'_k を次式により得られる $\bar{\mu}'_k$ に更新する。

$$\bar{\mu}'_k = \frac{\sum_i \sum_n (\omega'_i - \log n) p_n^k(\omega'_i) f'(\omega'_i)}{\sum_i \sum_n p_n^k(\omega'_i) f'(\omega'_i)} \quad (18)$$

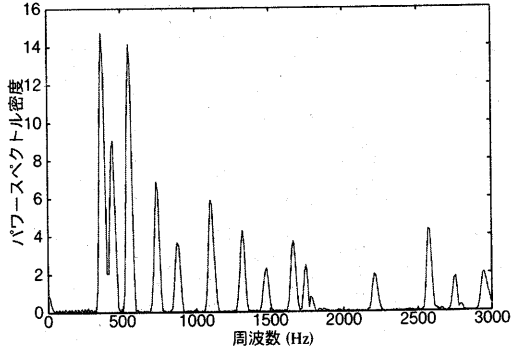


図 4 基本周波数が 371, 441, 556[Hz] の 3 音源多重音スペクトル

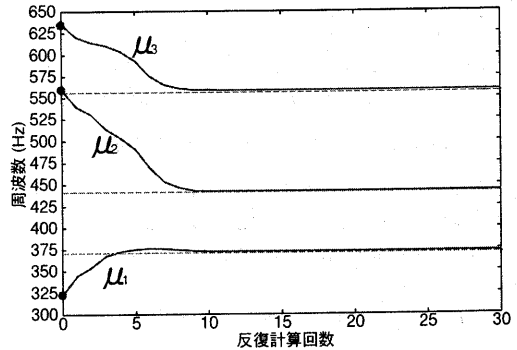


図 5 初期値が 323, 560, 635[Hz] のときの反復計算ごとの μ の更新値

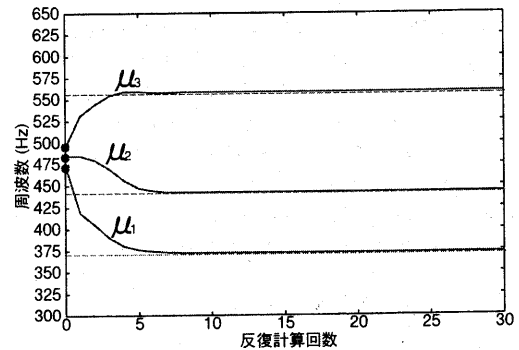


図 6 初期値が 474, 484, 495[Hz] のときの反復計算ごとの μ の更新値

ステップ 1 とステップ 2 の反復計算による $\mu = \exp(\mu')$ の収束値をスペクトル解析区間における推定基本周波数とする。

本アルゴリズムの収束の状況の例として、図 4 のスペクトルに対し、異なる例の初期設定の下での $\mu = \{\mu_1, \mu_2, \mu_3\}$ の反復計算ごとの更新値をそれぞれ図 5、図 6 に示す。ただし、点線は真の基本周波数を表す。

以上で述べた手法は、観測スペクトルの基本周波数成分のみのピークだけでなくすべての倍音成分も含めた形でピークを求めるアルゴリズムである。従って、基本周波数成分が倍音成

分と比べて極端に小さい音源であっても、先述した仮定から大きく外れるような音源でない限り推定に大きな問題はない。また、音源の音色を決定する要素の1つである倍音比のモデリングを必要としなくても推定が可能であることも大きな特徴である。ただし、次のようないくつかの課題となる点が挙げられる。

- (i) 収束値がクラスタ重心の初期値に大きく依存する。
- (ii) 音源数が既知である必要がある。
- (iii) クラスタ重心が、正しい基本周波数のオクターブ違いの値においても目的関数 $D_5(\mu')$ は極値をもつため、オクターブの判別は原理的に困難である。

5. 評価実験

提案手法“ハーモニック・クラスタリング”を用いてモノラル音楽音響信号を対象にピッチトラッキングの実験を行った。そこで、提案手法によるトラッキングの性能を、次の2点に着目して評価した。

- 倍音比が大きく変動する音源に対し、有効かどうか。
- ピッチが連続的に変動するとき、追従可能かどうか。

前者を確認するために歌唱、後者を確認するためにヴァイオリン演奏によるそれぞれの音楽音響信号を、ピッチトラッキングの実験対象として用いた。

5.1 ピッチトラッキングのための実験系

実験系の概要図を図7に示す。スペクトル解析部においては、サンプリング周波数 44.1[kHz]、フレーム長 50[ms]、フレームシフト 25[ms]とし、窓関数はハミング窓を用いてFFTを行い、スペクトル系列を得た。対象曲の開始フレームにおける各音源の基本周波数をそれぞれのクラスタ重心の初期値として与え、次のフレーム以降は前のフレームにおける推定基本周波数を初期値とした。また、クラスタ数は曲中の最大同時発音数とし、固定とした。実験に用いた対象曲の曲名、演奏方式、演奏者は表1の通りである。

表1 対象曲(実験データ)

作曲家、曲名	演奏方式	演奏者
滝廉太郎:“荒城の月”	独唱(無伴奏)	ソプラノ歌手
J. Pachelbel:“Kanon”	ヴァイオリン三重奏(通奏低音無し)	プロの演奏家と生徒2人

5.2 正解率の計算

音名正解率の計算方法として以下の処理を行った。まず、推定基本周波数系列を5フレームの移動平均処理により平滑化し、1フレームごとに最も近い音名に割り当てる。次に、このフレームごとの推定音名系列 $\mathbf{a} = \{a_1, \dots, a_s, \dots, a_S\}$ と、楽譜より得られる正解音名系列 $\mathbf{b} = \{b_1, \dots, b_t, \dots, b_T\}$ とのDPマッチングにより累積距離 $g(S, T)$ を計算する。 $g(s, t)$ は以下となる。 $\text{error}(s, t)$ は $a_s = b_t$ のときは0、 $a_s \neq b_t$ のときは1で与えられる。すなわち、挿入・脱落・置換のいずれかが1つあった場合に累積距離は1増加する。 $g(s, 0) = s$ 、 $g(0, t) = t$

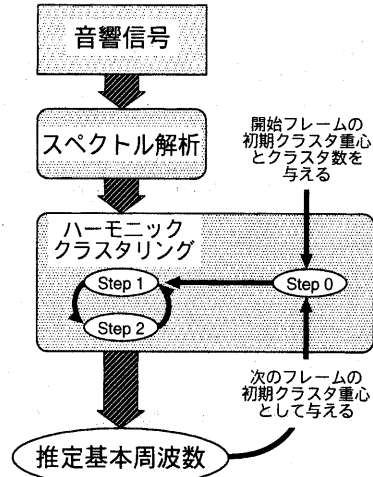


図7 ピッチトラッキングのための実験系

とし、式(19)により $g(S, T)$ まで順次求め、誤り数とした。

$$g(s, t) = \min \left\{ \begin{array}{l} g(s-1, t) + \text{error}(s, t) \\ g(s-1, t-1) + \text{error}(s, t) \\ g(s, t-1) + 1 \end{array} \right\} \quad (19)$$

従って、正解率は以下により得られる。

$$\text{accuracy}(\%) = \frac{S - g(S, T)}{S} \times 100 \quad (20)$$

5.3 実験結果

各実験データに対する音名正解率は表2の通りであった。また、独唱による“荒城の月”およびヴァイオリン三重奏による“Kanon”に対する推定結果の一部をそれぞれ楽譜とともに図8、図9に示す。破線は、楽譜により得られる正解音名に相当する基本周波数の系列を表す。図8には、推定基本周波数の系列に合わせて歌詞を記した。また、図9において丸で囲んだ部分はオクターブ誤りの例を表す。

表2 実験結果

作曲家、曲名	音源の種類	音源数	音名正解率(%)
滝廉太郎:“荒城の月”	歌声	1	90.6
J. Pachelbel:“Kanon”	ヴァイオリン	3	83.2

独唱による実験データに対する推定結果に見られた誤りは、ビブラートによるピッチ変動の範囲が上下 $\frac{1}{4}$ 音以上に及ぶことが多いこと、および過渡的にピッチが変化することなどが主な原因だった。これは、楽譜と実際の音響信号との間における相違によるものであり、本質的な推定誤りではないと考えられる。

ヴァイオリン三重奏による実験データに対する推定結果に関しては、最大同時発音数より発音数が少ないときに挿入誤りが見られた。例えば、クラスタ数が3、実際の発音数が2のとき、3つのクラスタ重心のうち2つは実際に発音されている基本周波数である目的値に収束するが、余ったクラスタ重心は目的値

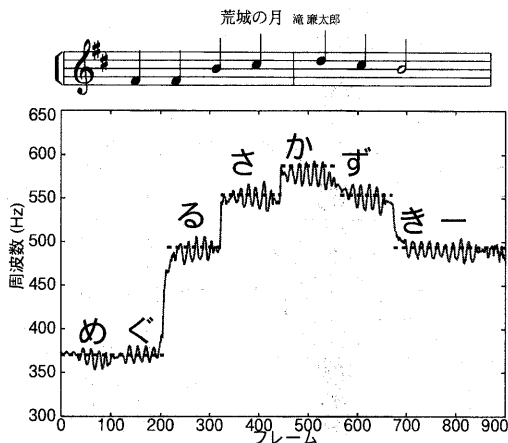


図8 “荒城の月”の楽譜の一部と、同部分のピッチトラッキング結果

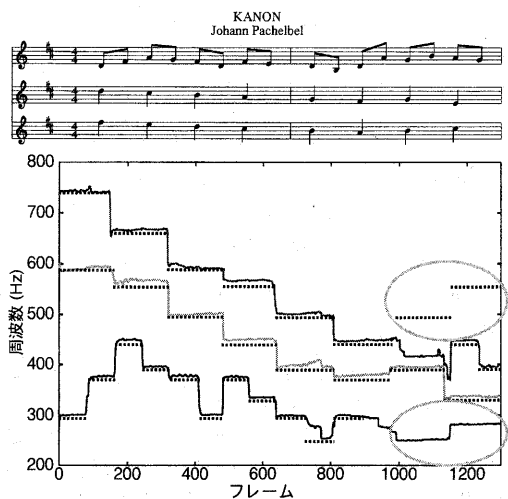


図9 “Kanon”の楽譜の一部と、同部分のピッチトラッキング結果

以外の局所解に収束する。これは、クラスタ数を固定しているため、発音数が固定ではない限り不可避な誤りと考えられる。

両実験データに対する全体の推定誤りのうち27%がオクターブ誤りであった。今回提案手法を適用したピッチトラッキングは、単純に初期値から最も近い極値を探索する程度のもので、4.で述べたようにオクターブ誤りを回避できなかった。ただし、これは提案手法の限界を意味するのではなく、今回のような単純なピッチトラッキング方法では対処しきれない問題であることを意味し、提案手法の適用手段により十分対処が可能である。

6. おわりに

本報告では、倍音比のモデリングを用いず、高い分解能で基本周波数を推定できる手法、“ハーモニック・クラスタリング”を提案した。

本手法を用いたピッチトラッキングの実験を行い、歌唱に対して90.6(%)、ヴァイオリンによる三重奏に対して83.2(%)の音名正解率を得た。結果より、倍音比変動の激しい歌唱に対し十分な性能が確認でき、倍音比のモデリングを必要としない提案手法の利点を示すことができた。また、歌唱およびヴァイオリンのビブラートをうまく追従できることが確認できた。ヴァイオリンの三重奏に対してもある程度の正解率を得ることができ、本手法の多重音に対する性能を確認できた。

ただし、初期値、音源数、オクターブに関する課題がまだ残されており、今後これらを解決する方法を検討する予定である。本報告においては、推定対象を調和性をもつ音源に限定していたが(仮定3)、倍音構造モデルは、必ずしも倍音周波数を基本周波数の整数倍とする必要はなく、非調和性をもつ楽器を対象とする場合は、その非調和性の構造をモデルとして適用すればよいので、対処が可能である。また今回は、クラスタ重心の推定しか行わなかったが、混合正規分布の重み、分散の推定も取り入れ、より洗練していくことが可能である。

7. 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。また、演奏データを提供された被験者の方々に感謝する。

文 献

- [1] 白土保: “二重奏音からの基本周波数分離抽出,” 日本音響学会誌, Vol. 54, No. 10, pp. 715-719, 1998.
- [2] 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.
- [3] 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, pp. 417-423, 1997.
- [4] 長束哲郎, 片寄晴弘, 井口征士: “異種楽器の音源分離システム,” 電子情報通信学会全国大会予稿集, D-5, 48, 1991.
- [5] 長束直樹, 才脇直樹, 井口征士: “異種楽器を対象とした採譜システム,” 電子情報通信学会'92春大会予稿集, D-499, 1992.
- [6] 三輪多恵子, 田所嘉昭, 斎藤賢: “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.
- [7] 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,” 電子情報通信学会論文誌, D-II, Vol. J84-D-II, No. 1, pp. 12-22, 2001.
- [8] M. Goto: “A Predominant-F0 Estimation Method for Real-world Musical Audio Signals: MAP Estimation for Incorporating Prior Knowledge about F0s and Tone Models,” Proceedings of CRAC-2001, 2001.
- [9] M. Goto: “A Predominant-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.
- [10] 亀岡弘和, 西本卓也, 篠田浩一, 嵯峨山茂樹: “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,” 日本音響学会 2003年春季研究発表会講演論文集, 3-7-3, pp. 837-838, 2003.