

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

亀岡 弘和 西本 卓也 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

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

あらまし 本報告では、多重音のモノラル音響信号の基本周波数抽出における音源数推定およびオクターブ位置推定を扱う。ある周波数におけるパワースペクトル密度は、ある一定の大きさの微小エネルギーを多数個加算したものと考え、その微小エネルギーに対し周波数軸上でのクラスタリングを適用する。まず単一音の場合、単純なクラスタリングではなく、倍音構造としての拘束をもつ複数のクラスタから成るクラスタ群の決定により基本周波数を抽出できる。このアプローチを我々はハーモニック・クラスタリングと呼ぶ。次に、微小エネルギーのクラスタ帰属度を正規分布に基づく確率的な尺度で扱い、同様な拘束条件をもつ複数のクラスタ群を同時に決定する問題を EM アルゴリズムにより解くことで、異なる音源同士で重なり合う倍音スペクトル成分を確率的に分解し、複数の基本周波数を抽出できる。我々はこれまでに、ハーモニック・クラスタリングを適用したピッチトラッキングのための実験系を構成し、収束性能の確認を行った。本報告では、これまで扱わなかった音源数およびオクターブ位置推定の問題を対象にし、クラスタリングの過程において AIC に基づいてこれらを推定する方法を検討する。これを取り入れた基本周波数抽出アルゴリズムの動作確認を、実演奏を録音したモノラル音響信号を対象とした実験により行う。

キーワード 多重音, 基本周波数抽出, k -means アルゴリズム, 混合正規分布, EM アルゴリズム, AIC

Estimation of Number of Sound Sources and Octave Position in Multi-Pitch Extraction Using Harmonic Clustering

Hirokazu KAMEOKA, Takuya NISHIMOTO, and Shigeki SAGAYAMA

Graduate School of Information Science and Technology, University of Tokyo

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

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

Abstract This report describes a method for estimating the number of sound sources and octave position in the process of extracting multiple fundamental frequencies of monaural audio signals. Considering that power spectral densities are sum total of a large number of micro-energies, clustering can be applied to the micro-energies in the frequency domain. We call this approach “Harmonic Clustering”. In case of single pitch, fundamental frequency can be estimated by determining a group of clusters which consist of multiple clusters having a constraint for constructing a harmonic structure. It is further extended to the case of multiple pitches by using a probabilistic measure based on Gaussian distributions for resolving the micro-energies to each cluster so as to separate overlapped harmonic spectral components and by giving multiple cluster groups each having the same constraint. Determining problem of the cluster groups can be formulated in a way similar to the EM algorithm. We have constructed a pitch tracking setup based on the Harmonic Clustering and have evaluated its convergence up to now. In this report, we focus on the problem of the number of sound sources and the octave position and propose methods for estimating them using AIC calculated during the process of the clustering. We experimented on extracting the fundamental frequencies of monaural musical audio signals by integrating these estimation methods stated above.

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

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呼ぶ。多重音の基本周波数推定は、音楽情報科学の分野において重要な要素の一つとして研究されている。効用としては、自動採譜、楽器音分離、演奏者の演奏分析、音響信号のMIDI圧縮などが考えられる。また、音楽情報科学以外にも、単一チャンネルの音源信号分離や雑音重畳下での音声認識に大きく貢献することが期待される。

従来より、音声の基本周波数推定法としてはケプストラム法や自己相関法などが用いられてきたが、音楽のように複数の基本周波数が混在する場合には適さなかった。そこで、音楽情報科学の分野において多重音または多声音楽を対象とした自動採譜の研究が1990年頃から盛んになった[1]～[3]が、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ことや、ミッシングファンダメンタル現象などが原因となり、音源数や各音源の基本周波数の推定は容易ではなかった。

柏野らは、和音構成情報、和音遷移情報、楽音ごとの倍音比情報などの統計データによる仮説を統合したベイジアンネットワークを構成し、事後確率が最大となる仮説の組(音名、音源名)、周波数成分を出力する手法を提案した[5],[6]。この手法では、観測パワースペクトルの極大点の集合の中から、各単音を構成する尤もらしい周波数成分を抽出することができるため、周波数解析における周波数分解能と同程度の分解能で基本周波数および倍音周波数を得ることができる。また後藤は、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を提案した[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、基本周波数を連続値として得ることができる。このように推定基本周波数値の分解能が高ければ高いほど、音源の分離合成技術や会話音声などのように非音楽的な基本周波数推定の精度向上などが期待できる。

上述の2手法では、音源の倍音比に関する情報を事前に仮定している。対象音源の倍音比情報が事前に得られるならば、それを積極的に活用することは原理的に有利である。しかし、事前に与えられない(未知の音源を対象とする)場合や、倍音比の仮定を置くこと自体が困難であるような(音声などのように時刻に応じてスペクトル包絡がさまざまに変化する)音源を対象とする場合などがある。

我々は、必ずしも音源の倍音比情報の仮定を必要とせずに単一チャンネルおよび単一フレームの短時間スペクトルからの基本周波数推定法“ハーモニック・クラスタリング”を提案し、これまでに簡易な実験系を構成し、このアルゴリズムの収束性の確認を行った[12],[13]。この実験系においては、音源数^(注1)が固定であるという仮定、および、真の基本周波数に収束するための初期値の許容範囲が十分広いという仮定の下で、直前フレームにおける収束値を初期値とする方法をとった。そのため、音

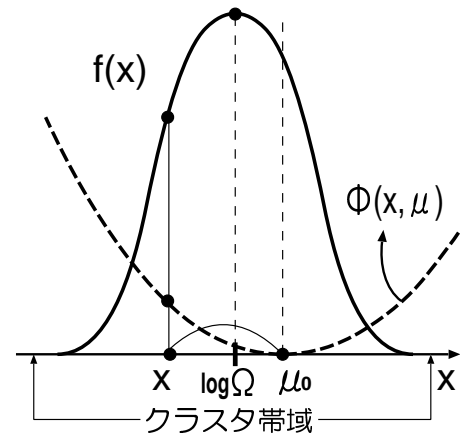


図1 純音スペクトル $f(x)$ と距離関数 $\varphi(x, \mu)$

源数が固定ではない場合やオクターブ位置の相違に関する扱いは考慮していなかった。そこで本報告では、AIC[11]を導入した音源数およびオクターブ位置の推定方法を検討し、これらを統合した基本周波数抽出アルゴリズムとしての性能を、モノラルの音楽音響信号を対象とした実験により評価する。次章では、ハーモニック・クラスタリングの原理と定式化について述べ、3章では音源数およびオクターブ位置推定方法について述べる。

2. クラスタリングとしての問題の定式化

2.1 線スペクトルの広がりクラスタ帯域の定義

短時間スペクトルの解析では、解析区間に窓関数を掛けることが一般的である。そのため、周波数 Ω が一定の単一正弦波(純音)の信号からでも、線スペクトルではなく、周波数 Ω 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。さらにピブラートのように、解析区間で周波数が連続的に変化する信号からは、それに応じた広がりをもつスペクトルが観測されることになるが、ここではこれを観測窓による広がりと同等に扱う。また、周波数を対数変換したとき、簡単のため、スペクトルの広がりには周波数に依存せず均一の幅をもつことにする。

パワースペクトル密度はある一定の大きさの微小エネルギーを多数個加算したたものであると考え、パワースペクトルは微小エネルギーの度数分布と解釈できる。これをクラスタ帯域と呼ぶことにする(図1)。

2.2 微小エネルギーのクラスタリング

周波数 ω 軸上に広がった各微小エネルギーに対しクラスタリングを適用する。対数周波数を $x = \log \omega$ 、周波数を対数変換したときの観測スペクトル密度を $f(x)$ と表すことにすると、クラスタ重心は次のような評価関数

$$D_{\text{pure}}(\mu) = \int_{-\infty}^{\infty} \varphi(x, \mu) \cdot f(x) dx \quad (1)$$

を最小化する μ を求めることで得られる(図1)。ただし、 $\varphi(x, \mu)$ を x と μ との距離を表す μ を中心に左右対称かつ下に凸の関数とする。($\varphi(x, \mu)$ が上に凸の関数の場合は、評価関数を最大化する μ を求める。) 得られたクラスタ重心は解析区間内での推定周波数 $\hat{\Omega}$ の対数である。

(注1): 本報告で言う音源数とは、同時発音する基本周波数の数をさす。

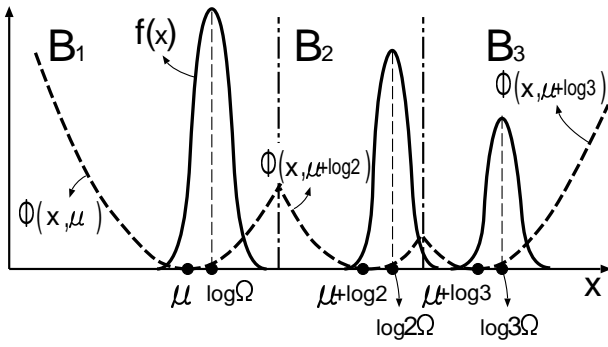


図 2 単一音スペクトルと距離関数およびクラスタ帯域

2.3 倍音構造の拘束を与えたクラスタリング

次に、倍音構造を持つ単一音信号の場合を考える．基本周波数，各倍音周波数それぞれの周辺の周波数幅をクラスタ帯域とし，複数のクラスタ重心を決定する．そこで，単純なクラスタリングではなく，それぞれのクラスタに倍音構造としての拘束を与え，自由度 1 のクラスタリングを行う．このような拘束をもつ複数のクラスタを倍音クラスタ群と呼びことにし，このクラスタリングを“ハーモニック・クラスタリング”と呼ぶ．

以後，倍音構造に関して倍音周波数は基本周波数の整数倍であると仮定する．すなわち，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が，これは単に以後の式を簡単にするためであり，整数倍の代わりに対象とする音源の倍音構造に合わせた拘束を与えれば同様に扱える．この仮定に従えば，クラスタ重心は $\mu, \mu + \log 2, \dots, \mu + \log n, \dots, \mu + \log N$ と置き，これらは基本周波数に相当する μ のみに依存することになる．この μ を基本クラスタ重心と呼ぶ．ただし，上限がナイキスト周波数の対数である範囲内にとりうるクラスタ重心の数を N とする．単一音の基本周波数は，評価関数

$$D_{\text{mono}}(\mu) = \sum_{n=1}^N \int_{B_n} \varphi(x, \mu + \log n) \cdot f(x) dx \quad (2)$$

を局所最小化する基本クラスタ重心 μ を求めることで得られる．ただし， B_n は第 n 倍音成分のクラスタ帯域とする． $\varphi(x, \mu + \log n)$ を x と $\mu + \log n$ とのユークリッド距離の二乗， B_n を図 2 のように $\mu + \log n$ と $\mu + \log(n-1)$ との中点および $\mu + \log n$ と $\mu + \log(n+1)$ との中点を両端とした x の帯域とすれば， k -means アルゴリズムとして定式化でき，クラスタ帯域の更新とクラスタ重心の更新を交互に繰り返し，収束させることでこの局所最小化問題を解くことができる．

2.4 多重音への拡張

上述した倍音クラスタ群の決定問題を，複数音源による多重音に対し，以下の 3 点に従って同様に扱う．

- (i) 複数 (K 個) の倍音クラスタ群を用いる．そこで，倍音クラスタ群 k の n 倍音帯域を B_n^k とし，その重心を $\mu_k + \log n$ とする．また，倍音クラスタ群 k において，上限がナイキスト周波数の対数である範囲内にとりうるクラスタ重心の数を $N(k)$ とする． μ を以下のようなベクトルとする．

$$\mu = \{\mu_1, \dots, \mu_k, \dots, \mu_K\} \quad (3)$$

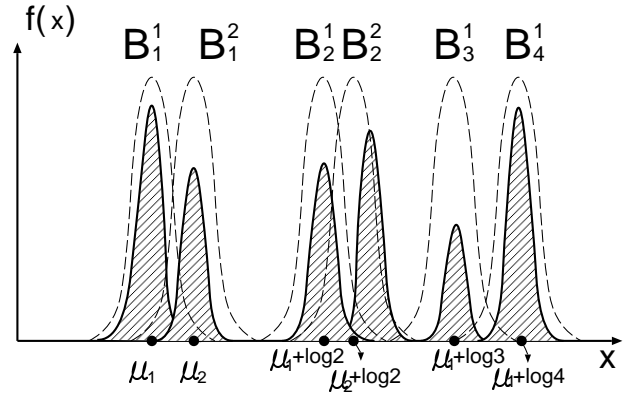


図 3 確率尺度によるクラスタ帰属度の導入

- (ii) それぞれのクラスタ帯域を確率的な尺度により分割する．そこで，周波数軸座標 x にある微小エネルギーのクラスタ帯域 B_n^k に対する帰属確率を $p_n^k(x)$ とする．ただし，以下を満たすものとする．

$$\sum_{k=1}^K \sum_{n=1}^{N(k)} p_n^k(x) = 1 \quad (4)$$

- (iii) クラスタごとの重み w_n^k を導入する． w を式 (5) のようなベクトルとする．ただし，式 (6) を満たすものとする．

$$w = \{(w_1^1, \dots, w_{N(k)}^1)^T, \dots, (w_1^K, \dots, w_{N(k)}^K)^T\} \quad (5)$$

$$\sum_{k=1}^K \sum_{n=1}^{N(k)} w_n^k = 1 \quad (6)$$

(ii) を行うのは，多重音の場合は異なる音源同士の基本周波数成分あるいは倍音周波数成分が重なり合って観測されるので，それらを分離することで，より厳密な解を得ることができると考えられるためである．また (iii) においてクラスタ重み w_n^k を導入するのは，音源間のパワー比や周波数成分間のエネルギー比 (倍音比) を考慮し，各クラスタに重要度を付加するためである．以上に従えば，評価関数は以下となる．

$$D_{\text{multi}}(\mu, w) = \sum_{k=1}^K \sum_{n=1}^{N(k)} \int_{-\infty}^{\infty} w_n^k \cdot \varphi(x, \mu_k + \log n) \cdot p_n^k(x) \cdot f(x) dx \quad (7)$$

ここで，観測スペクトルにおける広がり形状は，正規分布で近似できると仮定する．その場合，クラスタ帰属度 $p_n^k(x)$ は，平均 $\mu_k + \log n$ ，分散 σ の正規分布と (iii) で導入した重み w_n^k を用いて以下のような形式となる．

$$p_n^k(x) = \frac{w_n^k \cdot g(x|\mu_k + \log n, \sigma^2)}{\sum_k \sum_n w_n^k \cdot g(x|\mu_k + \log n, \sigma^2)} \quad (8)$$

$$g(x|\mu_k + \log n, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\{x - (\mu_k + \log n)\}^2}{2\sigma^2}\right\} \quad (9)$$

さらに，距離関数 $\varphi(x, \mu_k + \log n)$ を，以下のような $\mu_k + \log n$

を平均とし、上記と正規分布の x における対数尤度

$$\varphi(x, \mu_k + \log n) = \log \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{\{x - (\mu_k + \log n)\}^2}{2\sigma^2} \quad (10)$$

と置くことで、この評価関数の最大化問題は混合正規分布のパラメータの最尤推定と非常に類似した定式化となる。

2.5 複数の基本周波数の最尤推定アルゴリズム

上述のような評価関数を最大化するアルゴリズムを、混合正規分布のパラメータの最尤推定として広く用いられる EM アルゴリズムを応用することで導く。

平均 $\{\mu_k, \mu_k + \log 2, \dots, \mu_k + \log n, \dots\}$ 、分散 σ 、各重みが w の混合正規分布を K 個混合した分布を音源 K 個の多重音モデル $P_\theta(x)$ とし、このモデルのパラメータ θ は以下で与えられる。

$$\theta = \{\mu, w, \sigma(\text{固定})\} \quad (11)$$

先述したように $f(x)$ を微小エネルギーの度数密度関数と考えれば、 x について、パラメータ θ を $\bar{\theta}$ に更新したときのモデルの対数尤度の差は、

$$f(x) \log P_{\bar{\theta}}(x) - f(x) \log P_\theta(x) = f(x) \log \frac{P_{\bar{\theta}}(x)}{P_\theta(x)} \quad (12)$$

となる。Dempster らは、式 (12) において $f(x)$ を確率密度分布とし、平均対数尤度を最大にするために EM アルゴリズムを定式化した [10]、 $f(x)$ をスペクトル密度分布に置き換えても同様に定式化できる。 x がどの正規分布によって生成されたかは一意に決定できないため、これを直接最大化することはできない。そこで両辺に対し、 x がどの正規分布から生成されたかを表す $P_\theta(n, k|x)$ についての期待値をとり、この期待値を最大化するパラメータを求める。そこで、 Q 関数と呼ぶ以下のような評価関数

$$Q(\theta, \bar{\theta}) = \sum_{k=1}^K \sum_n \int_{-\infty}^{\infty} P_\theta(n, k|x) f(x) \log P_{\bar{\theta}}(x, n, k) dx \quad (13)$$

を置くことで、

$$\int_{-\infty}^{\infty} \left\{ f(x) \log P_{\bar{\theta}}(x) - f(x) \log P_\theta(x) \right\} dx \geq Q(\theta, \bar{\theta}) - Q(\theta, \theta) \quad (14)$$

が成立し、 Q 関数を最大化する $\bar{\theta}$ を求めることで、すべての x に対するモデルの対数尤度を単調に増加させることができる。 $P_\theta(n, k|x)$ は、

$$P_\theta(n, k|x) = \frac{P_\theta(x, n, k)}{P_\theta(x)} \quad (15)$$

$$= \frac{w_n^k \cdot g(x|n\mu_k, \sigma^2)}{\sum_k \sum_n w_n^k \cdot g(x|n\mu_k, \sigma^2)} \quad (16)$$

となり、式 (7) の $p_n^k(x)$ と一致する。また、 $\log P_{\bar{\theta}}(x, n, k)$ は、各正規分布が与える対数尤度なので、

$$\log P_{\bar{\theta}}(x, n, k) = \log \frac{w_n^k}{\sqrt{2\pi\sigma^2}} - \frac{\{x - (\mu_k + \log n)\}^2}{2\sigma^2} \quad (17)$$

であり、式 (7) の $\varphi(x, \mu_k + \log n)$ と一致する。従って、式 (7) の評価関数を最大化する μ, w を求めることと、多重音モデル $P_\theta(x)$ の最尤パラメータを求めることは同値であることが分かる。

以上より、初期設定 (ステップ 0) を経て、以下のようなステップ 1 とステップ 2 の反復計算の収束性は保証され、評価関数または Q 関数を局所最大化する μ, w を得ることができる。

ステップ 0: (初期設定)

クラスタ重心 μ 、クラスタ重み w の初期値を与える。

ステップ 1: ($p_n^k(x)$ の更新, E ステップに相当)

式 (7) により $p_n^k(x)$ を算出し、 $D_{\text{multi}}(\mu, w)$ を求める。

ステップ 2: (μ の更新, M ステップに相当)

$p_n^k(x)$ を固定の下で、

$$\{\mu, w\} = \underset{\{\bar{\mu}, \bar{w}\}}{\operatorname{argmax}} D_{\text{multi}}(\bar{\mu}, \bar{w}) \quad (18)$$

により μ, w を更新後、ステップ 1 に戻る。

3. 音源数とオクターブ位置の推定

真の基本周波数と同一のピッチクラス^(注2)の基本周波数においても評価関数 (7) は極大値をもつと考えられるため、上の反復計算により真の基本周波数を必ずしも得ることはできない。また、倍音クラスタ群の数が音源数と同数である必要があるが、対象とする多重音に含まれる音源数が不明である場合を想定すると、事前に与えることはできない。

そこで、この章では、AIC と呼ぶ情報量規準の概要を述べ、ハーモニック・クラスタリングと AIC により音源数およびオクターブ位置を推定する方法について述べる。

3.1 AIC によるモデルの選択基準

1 つのモデルに対する最尤パラメータ推定法は EM アルゴリズムにより求められるが、モデルの候補が複数個あるとき、その中から最良のモデルを選択する規準が必要となる。そこで、赤池によって提唱された AIC (Akaike Information Criterion, 赤池情報量規準) [11] を導入する。AIC は

$$\begin{aligned} \text{AIC} = & -2 \times (\text{モデルの最大対数尤度}) \\ & + 2 \times (\text{モデルの自由パラメータ数}) \end{aligned} \quad (19)$$

で与えられ、最小となるモデルを最良のモデルと考える。

3.2 処理行程 1: 音源数推定とクラスタ重心の局所探索

対象とする音響信号に含まれる音源数は不明なので、さまざまな倍音クラスタ群数の場合の多重音モデルが候補となる。そこで、多数の倍音クラスタ群から開始し、尤度に比較的大きく関与しないと見なせる倍音クラスタ群を順次削減させながら、その都度、各個数の下での多重音モデルが与える尤度を最大に

(注2): オクターブ位置を区別せずに分類された音高の集合

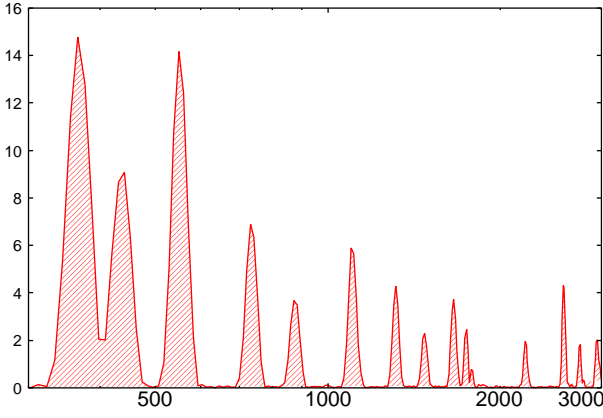


図 4 基本周波数が 371Hz, 441Hz, 556Hz の音が混在する多重音スペクトル (ヴァイオリン × 3)

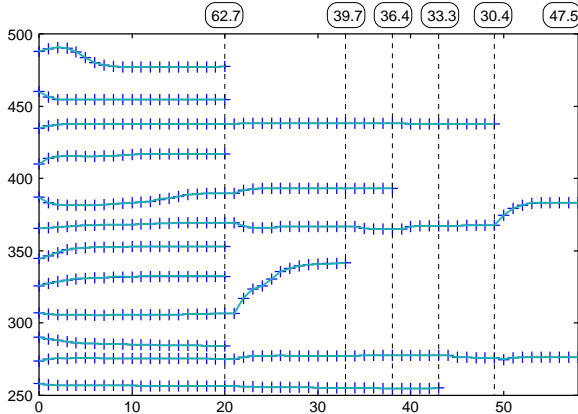


図 5 行程 1 による倍音クラスタ群数および基本クラスタ重心の更新

するパラメータ μ , w を EM アルゴリズムにより求め, AIC を計算する. AIC が上昇する前の倍音クラスタ群の数を推定音源数とする. 具体的な手順を以下に示す.

- (1) あるオクターブ範囲内における 12 平均律音階それぞれの基本周波数値を基本クラスタ重心の初期値とする. 従って倍音クラスタ群の初期の個数 K は 12 である.
- (2) ステップ 1, ステップ 2 の反復計算 (EM アルゴリズム) により最尤パラメータを求める. ただし, ここではクラスタごとの重み w_n^k に関して

$$w_1^k = w_2^k = \dots = w_{N(k)}^k (= w^k) \quad (20)$$

のような拘束を与える. これは, クラスタごとではなく, 倍音クラスタ群ごとの重みを後述する倍音クラスタ群の個数削減の規準に用いるためである. この場合, ステップ 2 における μ_k , w^k の更新値は以下で与えられる. ただし, $F = \int_{-\infty}^{\infty} f(x) dx$ とする.

$$\bar{\mu}_k = \frac{\sum_{n=1}^{N(k)} \int_{-\infty}^{\infty} (x - \log n) p_n^k(x) f(x) dx}{\sum_{n=1}^{N(k)} \int_{-\infty}^{\infty} p_n^k(x) f(x) dx} \quad (21)$$

$$\bar{w}^k = \frac{1}{FN(k)} \sum_{n=1}^{N(k)} \int_{-\infty}^{\infty} p_n^k(x) dx \quad (22)$$

- (3) AIC を算出する. 倍音クラスタ群ごとに 2 つの自由パラメータ μ_k , w^k があるので, 自由パラメータ総数は $2 \times K$ である. AIC が上昇した時点で終了し, 削減前の倍音クラスタ群数 \hat{K} を推定音源数とする.
- (4) 以下により倍音クラスタ群を削減し, 残った倍音クラスタ群数を \bar{K} とする. $K = \bar{K}$ とし, (2) に戻る.
 - (a) 重みが最小の倍音クラスタ群を消滅させる. これは, 重み最小の倍音クラスタ群はモデルに与える尤度に関する重要度が低いと見なせるためである.
 - (b) 2 つの基本クラスタ重心間の距離が $(1/12) \log 2$ より小さいとき, 重みの小さい方の倍音クラスタ群を消滅させる. これは, 上下 1/2 音の範囲内に他のクラスタ重心が存在するとき, 1 つの極値に 2 つのクラスタ重心が収束していると考えられるためである.

この行程を図 4 のスペクトルに対して実際に行った例を図 5 に示す. 破線が, (2) において収束判定により μ と w が収束したと見なされた時点を表し, 丸で囲った数値は (3) において計算した AIC の値である. 倍音クラスタ群の数が 2 のときに AIC が初めて上昇したため, この場合推定音源数は 3 となる.

3.3 処理行程 2: オクターブ位置推定

行程 1 により得られる基本クラスタ重心 μ_k に対し, オクターブを低い位置から順に上げていき, その都度, 各位置の下の多重音モデルが与える尤度を最大にするパラメータ w を EM アルゴリズムにより求め, AIC を計算する. AIC が上昇する前の段階におけるオクターブ位置を推定位置とする. オクターブ位置を表す指標 t (自然数) を用いて, 以下に示す手順を行程 1 における推定音源数 \hat{K} 個の倍音クラスタ群すべてについて行う.

- (1) t の初期値を 1 とする.
- (2) 倍音クラスタ群 k における基本クラスタ重心を $\mu_k + \log t$ と置く. そのとき, 上限がナイキスト周波数の対数である範囲内にとりうるクラスタ重心の数を $N(k, t)$ とする.
- (3) ステップ 1, ステップ 2 の反復計算 (EM アルゴリズム) により最尤パラメータを求める. ステップ 2 における w_n^k の更新値は以下で与えられる.

$$\bar{w}_n^k = \frac{1}{F} \int_{-\infty}^{\infty} p_n^k(x) dx \quad (23)$$

- (4) AIC を算出する. このとき, 自由パラメータ総数は $N(k, t)$ である. AIC が上昇した時点で終了し, 前の段階における t を推定オクターブ位置の指標とする. t を 1 増やし, (2) に戻る.

上述した 2 つの行程によって選択される多重音モデルは, AIC による規準の下で最良モデルとなる保証はないため, 上記した方法により音源数およびオクターブ位置がどの程度正しく推定できるか動作確認を行う必要がある.

表 1 対象曲 (実験データ)

作曲者, 曲名	演奏方式	演奏者
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

4. 動作実験

3章で示した2段階の処理行程を統合した基本周波数抽出アルゴリズムの動作実験を, モノラル音楽音響信号に対して行った. スペクトル解析は, サンプリング周波数を 44.1kHz, フレーム長を 25ms, フレームシフトを 10ms とし, Hamming 窓を窓関数として FFT (高速フーリエ変換) を行い, 短時間スペクトル系列を得た. 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.

4.1 音名正解率

性能評価の目安として以下のヒューリスティックな方法により音名正解率を与えた.

抽出した基本周波数を最も近い音名の基本周波数に割り当てる. 実験データの音響信号を 1 小節ごとに分割し, 各小節のフレーム数を得る. 楽譜に記譜されている音符に応じて各音の音長フレーム数 (四分音符の場合は 1 小節のフレーム数の 1/4) を与え, 音長フレーム数の 1/2 以上のフレーム数だけ正解音名に相当する基本周波数を抽出できた場合に, 1 つの音に対して正解とした. また, 同一のピッチクラスで, オクターブ位置の異なる複数の音が同時発音している場合には, オクターブ位置の低い方の基本周波数を抽出できていれば, 正解とした. 正解した数を以上の基準に従って目視により数え, 総音符数を分母として正解率を計算した.

4.2 実験結果

ヴァイオリン三重奏による "Kanon" に対する基本周波数の抽出結果の一部をそれぞれ楽譜とともに図 6 に示す. ×印がフレームごとに抽出した基本周波数値を, 破線が正解音名に相当する基本周波数を表す. 音名正解率は, 92.7%であった. 図 6 の通り, 全体として音源数およびオクターブ位置に関して妥当な推定結果が得られていることが分かる.

5. おわりに

本報告では, ハーモニック・クラスタリングと AIC の導入による音源数の推定法とオクターブ位置の推定法を統合し, 単一チャンネルおよび単一フレームの短時間スペクトルから基本周波数を抽出するアルゴリズムを構成し, モノラル音楽信号を対象としてアルゴリズムの動作確認を行った. 今後, より多くのデータに対して実験を行い, 詳細な評価を行う必要がある.

本アルゴリズムは対象音源の倍音比情報を必要としない利点はあるが, その反面, オクターブ位置の異なる同一ピッチクラスの同時発音に対し, それぞれの基本周波数を抽出することはできない. 従って今後もこの利点を保持した方法論を追求していく場合には, 時間推移情報を取り入れたり, 倍音比情報を自動獲得する方法など, さらなる検討を行いたい. 一方, 不完全ながらも倍音比に関する仮定や情報が利用できる場合もあり, このような場合の扱いに関しても考慮していきたい.

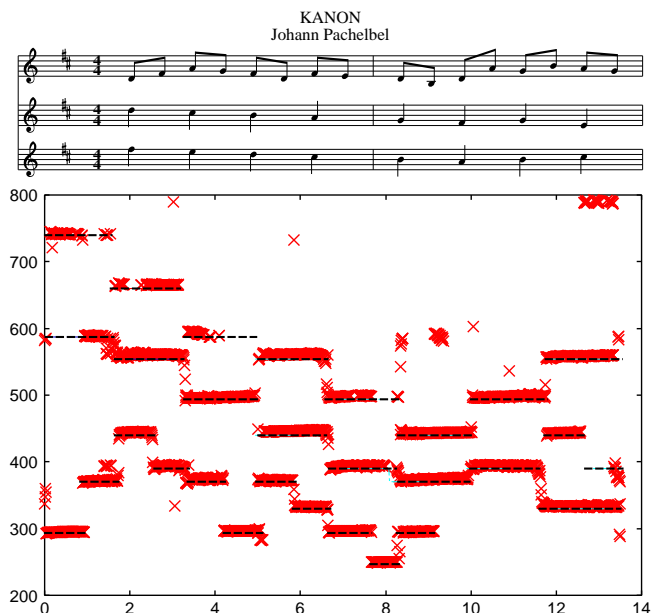


図 6 "Kanon" の楽譜の一部と同部分の基本周波数抽出結果

文 献

- [1] 白戸保: "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] 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.
- [5] 柏野邦夫, 中臺一博, 木下智義, 田中英彦: "音楽情景分析の処理モデル OPTIMA における単音の認識," 電子情報通信学会論文誌, D-II, Vol. J79-D-II, No. 11, pp. 1751-1761, 1996.
- [6] 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.
- [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] A. P. Dempster, N. M. Laird and D. B. Rubin: "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. of Royal Statistical Society Series B*, Vol. 39, pp. 1-38, 1977.
- [11] H. Akaike: "Information Theory and an Extension of the Maximum Likelihood Principle," *2nd Inter. Symp. on Information Theory*, Akademia Kiado, Budapest, pp. 267-281, 1973.
- [12] 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.
- [13] 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アルゴリズム," 情報処理学会研究報告, SIGMUS50-5, pp. 37-43, 2003.