Vol.2011-MUS-92 No.1
2011/10/11

自譜めくりシステムのための多重音の音高・音源数の推定法の高速化

安部翔[†]　小田弘良[†]　松島俊明[†]

我々は自動譜めくりシステムの開発を進めており，そのために多重音の音高・音源数の高速推定の研究を行っている．
以前，複素スペクトル内挿法により基本周波数候補を求め，それらの組み合わせによる同時発音数を評価する尺度を導入し，この評価尺度が最小なる音高の組み合わせを直接求める方法で，音源数が未知の音響信号からの高速推定法について報告したが，検出可能な同時発音数を多くするに従って計算量も増え，処理落ちが時々生じるという問題点があった．
そこで，今回，この評価尺度の最小値の変化率の推移に着目し，これを用いて多重音の音高・音源数の高速推定を試みた結果，以前よりも高速な推定が可能となったので報告する．

A fast multi-pitch estimation method for the automatic page turner

Syou Abe[†], Hiroyoshi Oda[†] and Toshiaki Matsushima[†]

We are developing an automatic page turner system, therefore are studying high-speed estimation of the number of notes and their pitches simultaneously from acoustic sound.
We already reported the following method: that is, to calculate the candidates of fundamental frequencies depending on the complex-spectrum interpolation, to introduce a evaluation function to search for the combination of the pitches which gives the minimum value of the evaluation function directly, and to presume the number of notes and their pitches from acoustic sound to a high speed.
However, by this method, computational complexity also increased as the number of detective notes increased, and there was a problem that processing omission sometimes occurred.
By considering transition of the rate change of the minimum value of the evaluation function, faster presumption of the number of notes and their pitches from acoustic sound can be attained than before.

1. はじめに

筆者らは以前より尺八譜の情報処理システムの研究を行っており，その機能の一部として自動譜めくりの実装を試みてきた[1]．自動譜めくり機能に要求される要素技術の１つに実時間での音高推定がある．尺八演奏だけに限れば単音を対象とすれば十分であるが，箏や三味線等などとの合奏曲への応用を考慮すると，複数音への対応も必要となる．

近年，複数の楽音を含んだ音響信号から音高情報を抽出する研究が盛んに行われるようになり，例えば混合正規分布モデルのパラメータ推定[2]や混合正規分布の調波時間構造のクラスタリング[3]などの方法が提案されており，目覚しい成果を挙げている．しかし，これらの方法では基本的にＥＭアルゴリズム等などによる繰り返し計算を用いているため，処理時間のコストが非常に大きくなってしまうため，自動譜めくりなど実時間処理が必要な用途に対しては，これらの方法を直接適用することは非常に困難である．

そこで，筆者らは，複素スペクトル内挿法[4]により求めた基本周波数候補と同時発音数の組み合わせを評価する関数を用い，この評価関数が最小となる推定音高の組み合わせを直接求めることで、音源数が未知な音響信号から音高・音数の高速推定を行う方法について既に報告した[5]．しかし，この方法では評価関数に現れるペナルティ係数の決定方法の問題や，推定可能最大音源数を増やした場合に処理落ちが生じるという問題があった．

今回，評価関数の値の減少傾向と元の音響信号に実際に含まれている音数の間に関連性があることに着目し，評価関数の値の推移を観測することでより高速な音高・音源数の同時推定を試みたので報告する．本報告では，既に報告済みの高速推定方法についてその概略の説明を行った後，今回，新たに考案した推定方法について述べ，それらの方法による実験結果を示す．

2. 音高推定のための評価関数

多重音の音高推定を行うための定式化方法はいくつかあるが,本研究では嵯峨山らにより考案されたハーモニッククラスタリング[6，7]における定式化と同様の方法により定式化を行う．音源に含まれる単音の基本周波数を Ω とすると，一般的に窓関数の影響等により観測されるスペクトルの形状はΩを中心として左右に単調減少しかつ対象な形状となる．一般の楽音の場合は倍音構造を持つため，倍音の周波数が基本周

[†] 東邦大学理学部情報科学科
Dept. of Information Sciences, TOHO University

波数 Ω の整数倍となることを仮定すると，ω を周波数 $f(\omega)$を観測スペクトルとして，

$$D(\mu)=\sum_n \int_{T_n} \varphi(\omega, n\mu)\cdot f(\omega)d\omega \quad (1)$$

ただし $$T_n=\left\{\omega \middle| n=\arg\min_m |\omega - m\mu|\right\} \quad (2)$$

を最小とする μ として Ω は与えられる．ここで，$\phi(\omega, n\mu)$は ω μ との距離尺度を表す関数で，$\omega = n\mu$ で極小値を取る左右対称な関数であれば何でもよいが，文献[6]と同様に

$$\varphi(\omega, n\mu)=(\omega - n\mu)^2 \quad (3)$$

に選ぶと，式(1)は

$$D(\mu)=\sum_n \int_{T_n} (\omega - n\mu)^2 \cdot f(\omega)d\omega \quad (4)$$

となる．またT_n は$(n-1)\mu$ と $n\mu$ の中点および$n\mu$ と $(n+1)\mu$ の中点を両端とする帯域を表している．

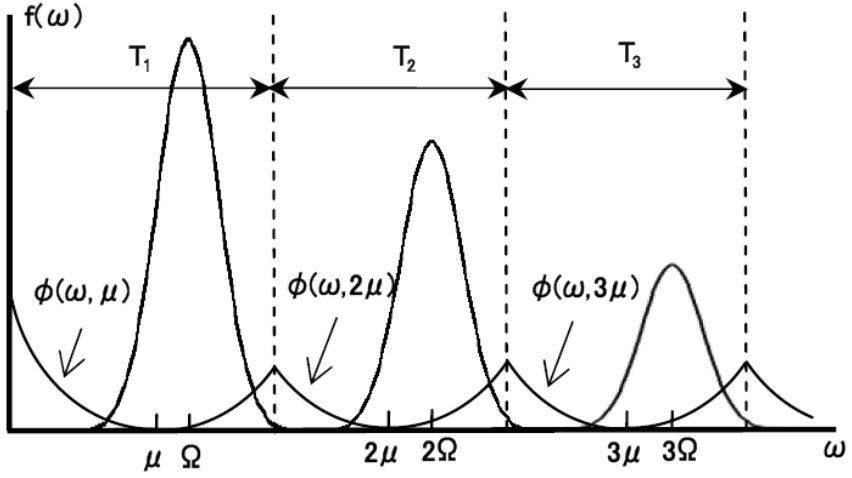

図 1　単音の場合の基本周波数の推定方法

音源数が複数の場合は，式(1)，(2)を複数音に拡張した評価関数，即ち，音源数を K，基本周波数の推定値を $\boldsymbol{\mu}=\{\mu_1, \mu_2, \cdots, \mu_K\}$とすると，評価関数

$$D(\boldsymbol{\mu})=\sum_{k=1}^{K}\sum_n \int_{T_n^k} \varphi(\omega, n\mu_k)\cdot f(\omega)d\omega \quad (5)$$

ただし $$T_n^k=\left\{\omega \middle| (n,k)=\arg\min_{m,l} |\omega - m\mu_l|\right\} \quad (6)$$

を最小とする $\boldsymbol{\mu}$ を求めれば良い．この最小値は，式(6)により与えられる周波数帯域 $T^k=T_1^k \cup T_2^k \cup \cdots$をクラスタと見なしてクラスタ重心である μ_k を

$$\mu_k=\frac{\sum_n n \sum_{\omega_i \in T_n^k} \omega_i f(\omega_i)}{\sum_n n^2 \sum_{\omega_i \in T_n^k} f(\omega_i)} \quad (7)$$

により更新するという反復計算を行い，その収束値を求めることで得ることができるが，収束するまでの反復処理に時間がかかる，収束値が初期値に依存し，誤った値に収束することが多々あるため様々な初期値に対して試行する必要がある，等の問題がある．そのため，クラスタリングによる解法を譜めくり等リアルタイム性が要求される用途に適用するのは現状では難しい．

そこで，本研究では，周波数ピークの検出精度に優れた複素スペクトル内挿法により検出されたピーク位置の中に正しい基本周波数が含まれている確率が非常に高い点に着目し，クラスタリングにより $D(\boldsymbol{\mu})$の収束値を求めるのではなく，様々な組み合わせの $\boldsymbol{\mu}$ に対して直接 $D(\boldsymbol{\mu})$の値を計算して $\boldsymbol{\mu}$ の推定を行うことにした．

3. 音高・音源数の同時推定アルゴリズム(音源数既知)

まず，音源数(和音数)K が既知の場合の音高推定アルゴリズムを示す．短時間フーリエ変換(STFT)により得られたパワースペクトルのうち，一定以上のパワーを持つ上位 L 個(ただし L > K とする)のピーク周波数 $\{\mu_1, \mu_2, \cdots \mu_L\}$を複素スペクトル内挿法により抽出する．この L 個の基本周波数候補から選び出した K 個の基本周波数を小さい順に並べたベクトル

$$\boldsymbol{\mu}=(\mu_{j(1)}, \mu_{j(2)}, \cdots, \mu_{j(K)}),\ j(p)=1,2,\cdots,L \quad (8)$$

に対して $D(\boldsymbol{\mu})$を算出し，最小の $D(\boldsymbol{\mu})$を与える $\boldsymbol{\mu}$ を基本周波数の推定値とする．従って

音源数 K が既知の場合のアルゴリズムは以下のようになる．

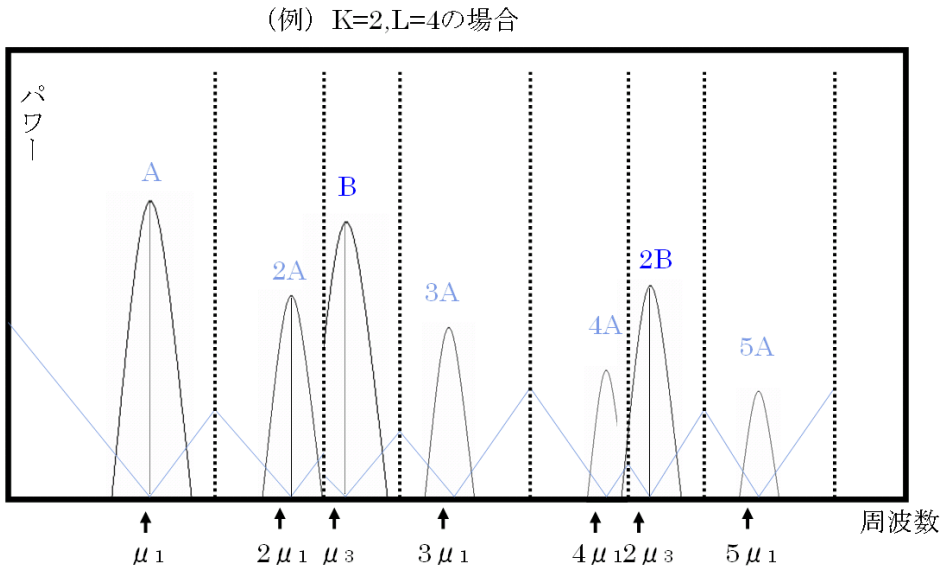

図 2．多重音に対するハーモニッククラスタリングの使用例

アルゴリズム1：音源数Kが既知の場合

① 複素スペクトル内挿法によりパワーのピーク位置を検出し，パワーの大きい順にL個の基本周波数候補 $\{\mu_1,\mu_2,\cdots\mu_L\}$ を選出する．

② L個の候補からK個を選択して得られる ${}_LC_K$ 個の基本周波数の組み合わせの集合を $M=\{\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\cdots,\boldsymbol{\mu}_{{}_LC_K}\}$ とする．

③ M の各要素について $D(\boldsymbol{\mu}_i)$ を算出し，最小値を取る $\boldsymbol{\mu}_i$ を基本周波数の推定値とする．即ち

$$\boldsymbol{\mu}=\arg\min_{\boldsymbol{\mu}_i} D(\boldsymbol{\mu}_i) \qquad (\boldsymbol{\mu}_i \in M) \qquad (9)$$

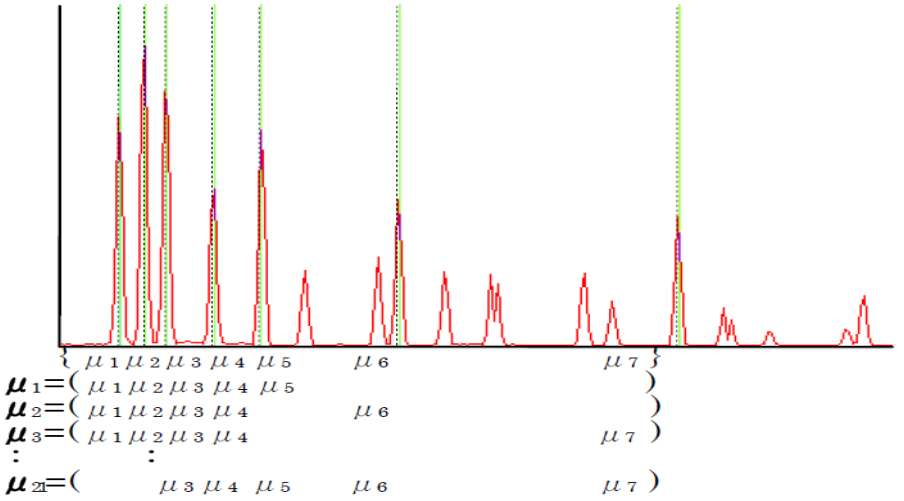

図 3．複素スペクトル内挿法によるピーク検出と基本周波数候補の選択例

4. 音高・音源数の同時推定アルゴリズム(音源数未知)

次に，音源数が未知の場合であるが，音源数が K の時の $\boldsymbol{\mu}$ を $\boldsymbol{\mu}^K$ と表すと，音源数が K の場合の式(5)の最小値 $\min D(\boldsymbol{\mu}^K)$ を K=1 から求めていき，最も小さい値を与える K および μ に決定する方法が考えられる．しかし，評価関数の値 $D(\boldsymbol{\mu}^K)$ は，一般的に $\boldsymbol{\mu}$ の次元 K に対して単調に減少していくので，実際の音源数よりも大きい次元の $\boldsymbol{\mu}$ で最小となることが分かる．そこで，$\boldsymbol{\mu}$ の次元数の増加に対するペナルティ係数 $p(K)$ を追加し，$p(K)D(\boldsymbol{\mu}^K)$ を新たな評価関数とする．従って最大検出音源数(検出可能な最大の同時発音数)を N とすると，音源数が未知の場合の音高推定アルゴリズムは以下のようになる．

アルゴリズム２：音源数 K が未知の場合

①複素スペクトル内挿法によりパワーのピーク位置を検出する．

②K＝１とする．

③パワーの大きい順に L 個（L＞K）の基本周波数候補 $\{\mu_1,\mu_2,\cdots\mu_L\}$ を選出す

Vol.2011-MUS-92 No.1
2011/10/11

る．

④L 個の候補から K 個を選択して得られる ${}_L C_K$ 個の基本周波数の組み合わせの集合を $M^K = \left\{ \boldsymbol{\mu}_i^K \middle| i = 1, 2, \cdots, {}_L C_K \right\}$ とする．

⑤ M^K の各要素について $D\left(\boldsymbol{\mu}_i^K\right)$ を算出し，最小値 $\boldsymbol{\mu}_i^K$ をとる音源数が K と仮定したときの基本周波数の推定値とする．即ち

$$\boldsymbol{\mu}^K = \arg\min_{\boldsymbol{\mu}_i^K} D\left(\boldsymbol{\mu}_i^K\right) \quad \left(\boldsymbol{\mu}_i^K \in M^K\right) \tag{10}$$

⑥K<N ならば K←K＋１として③へ戻る

⑦ $p(K)D\left(\boldsymbol{\mu}^K\right)$ を最小とする $\boldsymbol{\mu}^K$ を音源数 K と基本周波数 $\boldsymbol{\mu}$ の推定値とする．即ち

$$\left(\boldsymbol{\mu}, \mathrm{K}\right) = \arg\min_{\boldsymbol{\mu}^K} p(K) D(\boldsymbol{\mu}^K) \quad (K = 1, 2, \cdots, \mathrm{N}) \tag{11}$$

本アルゴリズムにより，音高・音源数を同時に推定することが可能であるが，実際の音源数 K の数に関係なく，検出可能最大同時発音数 N まで $D\left(\boldsymbol{\mu}_i^K\right)$ の値を計算する必要があるため，N の値が大きくなると処理速度が低下して処理落ちが生じ易くなる．また，最適なペナルティ係数 $p(K)$ を決定方法が難しい，という問題がある．これらの問題を解決する方法として，$\boldsymbol{\mu}^K$ の値の推移に着目した方法を次に述べる．

5. 音高・音源数の同時推定アルゴリズム(音源数未知・改良型)

式(5)からも分かるように，$D(\boldsymbol{\mu})$ の値は，$\boldsymbol{\mu}$ が基本周波数のみで構成されている場合には小さな値を取る．一方，基本周波数以外に加え，倍音の要素が加わっている場合には，クラスタ範囲が重複するだけなので，$D(\boldsymbol{\mu})$ の値を減少させる効果は大きくないと予想できる．即ち，$\boldsymbol{\mu}$ の要素を増やしていく場合，$\boldsymbol{\mu}$ に基本周波数の要素が追加されている間は $D(\boldsymbol{\mu})$ は減少を続けるが，すべての基本周波数が $\boldsymbol{\mu}$ に含まれた後，倍音の要素が追加されるようになると，$D(\boldsymbol{\mu})$ の値の変化が小さくなることが期待できる．

この仮定が成り立っているかを実際に調べてみたところ，例えば３和音の場合は図４に示すように，推定音源数 K が実際の音源数に一致するまでは $D(\boldsymbol{\mu})$ の値は減少していくが，音源数よりも K が大きくなると $D(\boldsymbol{\mu})$ の値はほとんど変化しなくなる場合が多かった．

このことから，$D(\boldsymbol{\mu})$ の値の変化の様子，即ち減少傾向を調べ，減少幅が小さくなった時点で推定音源数を決定することができる．即ち，$D(\boldsymbol{\mu})$ の変化率 $r(K)$ を

$$r(K) = D\left(\boldsymbol{\mu}^{K+1}\right) \Big/ D\left(\boldsymbol{\mu}^K\right) \tag{12}$$

と定義すると，変化率が１に近いある一定値以上の値となった場合，即ち

$$r(K) > \alpha : \text{ただし } \alpha \text{ は１に近い定数} \tag{13}$$

を始めて満たした時の K は，実際の音源数より１つ大きい値を取ることになる．このことから，以下のアルゴリズムにより，音高・音源数の同時推定が可能となる．

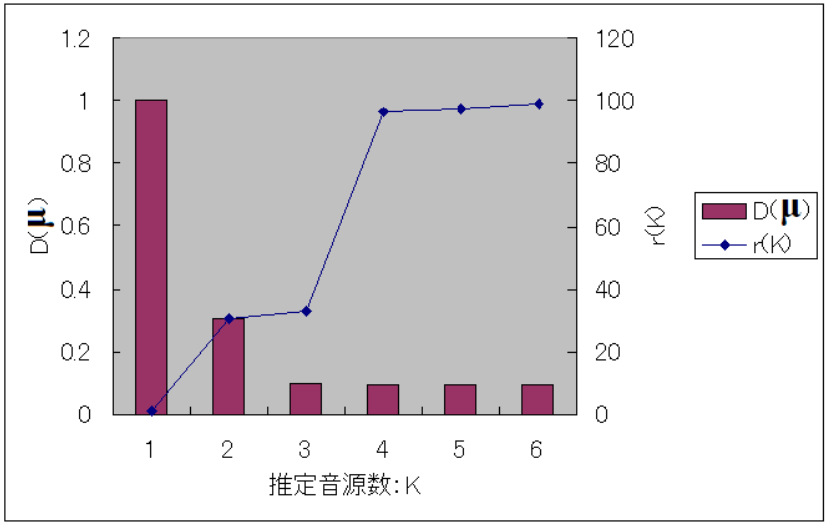

図 4. 推定音源数を増やしていった場合の D(μ)と r(K)の算出結果の推移

アルゴリズム３：音源数 K が未知の場合(改良型)

①複素スペクトル内挿法によりパワーのピーク位置を検出する．

②推定音源数 K＝１とする．

③パワーの大きい順に L 個（L＞K）の基本周波数候補 $\{\mu_1, \mu_2, \cdots \mu_L\}$ 選出する．

④L 個の候補から K 個を選択して得られる ${}_L C_K$ 個の基本周波数の組み合わせの集合を $M^K = \left\{ \boldsymbol{\mu}_i^K \middle| i = 1, 2, \cdots, {}_L C_K \right\}$ とする．

⑤ M^K の各要素について $D\left(\boldsymbol{\mu}_i^K\right)$ を算出し，最小値 $\boldsymbol{\mu}_i^K$ をとる音源数が K と仮定し

たときの基本周波数の推定値 $\boldsymbol{\mu}^K$ とする．即ち

$$\boldsymbol{\mu}^K = \arg\min_{\boldsymbol{\mu}_i^K} D(\boldsymbol{\mu}_i^K) \qquad (\boldsymbol{\mu}_i^K \in M^K) \tag{14}$$

⑥Ｋ＝１ならばＫ←Ｋ＋１として③に戻る。
⑦２≦Ｋ＜Ｎのとき，式(13)が成り立っていれば推定音源数をＫ－１に決定し，$\boldsymbol{\mu}^{K-1}$ を推定音高とする．そうでなければＫ←Ｋ＋１として③へもどる．
⑧Ｋ＝Ｎならば推定音源数をＫとし $\boldsymbol{\mu}^K$ を推定音高とする．

表１にアルゴリズム２とアルゴリズム３の $D(\boldsymbol{\mu})$ の計算回数の比較を示した．この表からも分かるように，アルゴリズム３では推定音源数を実際の音源数＋１まで計算すれば良いため，特に実際の音源数が少ない場合にアルゴリズム２に比べて計算回数を大幅に減らすことができることがわかる．またアルゴリズム２では必要となるペナルティ係数の設定が不要なため，アルゴリズム２の問題点を同時に解決することができる．

実際の音源数	アルゴリズム2	アルゴリズム3
単音	83	9
2和音	83	19
3和音	83	34
4和音	83	55
5和音	83	83
6和音	83	83

表１．　アルゴリズム２とアルゴリズム３の D($\boldsymbol{\mu}$)の計算回数の比較(L=K+2 とした場合)

6.　実験結果

以下に，提案アルゴリズムによる音高・音源数の推定実験の結果を示す．いずれも STFT の窓長は 4096，窓関数はハミング窓を用い，基本周波数候補数は N=6, L=K+2 とした．予備実験の結果から，アルゴリズム２におけるペナルティ係数の値は $p(K)=K$ を，またアルゴリズム３における変化率の閾値 α =0.8 を用いている．

まず，アルゴリズム２のシステムとアルゴリズム３のシステムの性能を比べるために MIDI 音源により作成した音源数が変化する音響データについて音源数・音高推定を行った結果を図５に示す(実験 1)．

アルゴリズム２に比べてアルゴリズム３では，ＣＤＥを含む和音のように，比較的基本周波数が近い音を含む和音で検出誤りが増えていることが分かる．しかし処理速度に関しては，単音から４和音に対してはアルゴリズム２よりもアルゴリズム３の方が処理速度は向上し，処理落ちの回数が明らかに少なくなった．

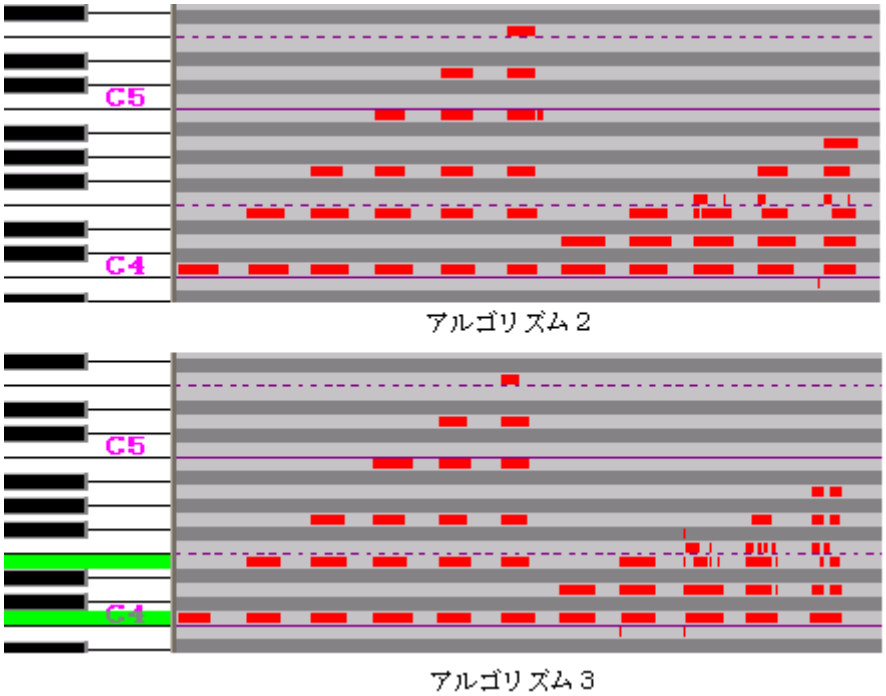

図 5．音源数が未知の場合の採譜結果
(アルゴリズム２と３の比較)

次に，RWC 研究用音楽データベース[8]に含まれている楽曲サンプル(MDB-J-2001 No.9)に対しての実験結果を図 6 に示す(実験２)．実験条件は実験１と同一である．どちらもやや検出もれが目立つが，和音を含む多くの音符が検出できており，音源数・音高の同時推定にある程度成功していることが分かる．処理落ちに関しては，アルゴリズム２は３和音程度から顕著になってきたのに比べ，アルゴリズム３では５和音以上で処理落ちが目立つようになった．アルゴリズム２に比べて処理速度の大幅な向上を実現できたと言えるが，まだまだ改善の余地が残されている．

Vol.2011-MUS-92 No.1
2011/10/11

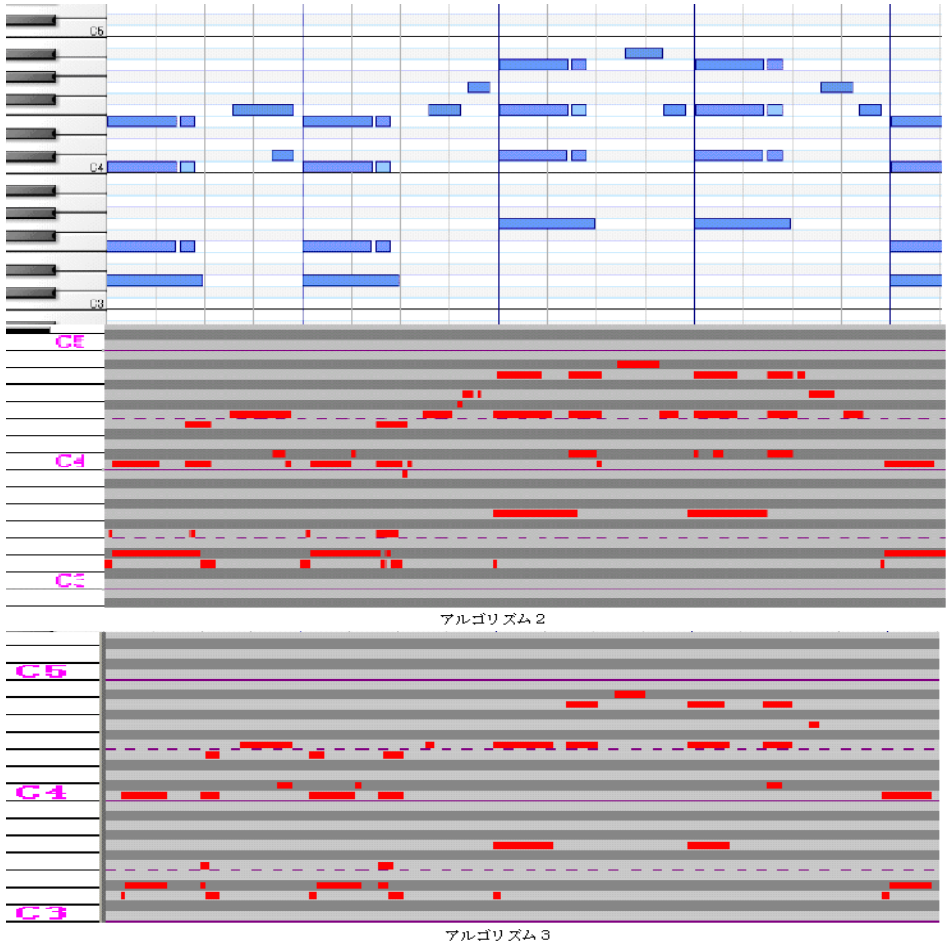

図 6. 実験３：楽曲サンプルデータに対する音高推定結果
(上段：正解 MIDI データ，中段：アルゴリズム２　下段：アルゴリズム３)

7. まとめ

一般的に音高と音源数の同時推定は難しい問題であるが，本方式では比較的簡単な処理により，高速に音高と音源数の同時推定を行うことができた．従来のアルゴリズム２では理想的なペナルティ係数の設定が難しかったが，今回改良したアルゴリズム３ではそのペナルティ係数を必要とせずに音高と音源数の同時推定が可能になった．しかし，実音を対象とした実験では，実際の音源数よりも検出できた同時発音数が少ないことが多いため，今後さらなる改善が必要である．

また，音源数が多い場合(概ね５和音以上)については，従来手法と処理量が変わらないため，処理落ちの問題は完全には解決できていない．また，低音部やオクターブ重なりの和音の検出精度が良くない点など，未解決の問題点が多く残されている．

しかし，自動譜めくりシステムでは，通常の採譜システムとは異なり，正解楽譜の情報をシステムが持っている状態での処理であるため，必ずしもすべての音符情報を正確に認識できる必要はない．４音程度までの和音の検出が可能であれば，既知の楽譜中からの演奏箇所の推定は可能と思われるので，自動譜めくりシステムへの利用に十分適用可能な方法であると考えている．

参考文献

[1] 松島俊明，”尺八譜のマルチメディア情報処理”，画像電子学会第 34 回年次大会予稿集，pp.23-30(2006)
[2] 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，Vol.J84-D-Ⅱ，No,1，pp.12-22(2001)
[3] 亀岡弘和，西本卓也，嵯峨山茂樹，“調波時間構造クラスタリング(HTC)による音楽音響特徴量の同時推定”,情報処理学会研究報告，2005-MUS-61-12,pp.71-78(2005)
[4] 原裕一郎，井口征士，“複素スペクトルを用いた周波数同定”，計測自動制御学会論文集，Vol.19,No.9,pp.718-723(1983)
[5] 宮坂広純，小田良弘，松島俊明，“多重音の基本周波数評価尺度の最小値選択による実時間・音源数推定”，第７回情報科学技術フォーラム(FIT2008)，E-045，pp. 241-243(2008)
[6] 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の基本周波数推定アルゴリズム”，情報処理学会研究報告書，2003-MU-50-5，pp27-32(2003)
[7] 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と情報量基準による音楽の音高/音源数の推定”，情報処理学会研究報告，2005-MUS-62-5,pp.27-32(2005)
[8] 後藤真孝，橋口博樹，西村拓一，岡隆一，“RWC 研究者用音楽データベース：クラシック音楽データベースとジャス音楽データベース”情報処理学会研究報告，2002-MUS-44-5,pp.255-32(2002)