

自動譜めくりシステムのための多重音に対する 音高・音源数の高速推定法

安部翔 小田弘良 松島俊明

東邦大学 理学部 情報科学科

1. はじめに

筆者らは以前から自動譜めくりの実装を試みてきた[1]. 自動譜めくりに要求される要素技術の一つに実時間で音高推定がある.

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ってきた. 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, 目覚ましい成果を上げていが, これらの方法では基本的に EM アルゴリズム等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ったため, 自動譜めくりと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・音源数の高速推定を行った[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.

そこで, 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源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[6]. 本報告では, さらに明らかな倍音ピークを除去した方法を試み, それらの方法の処理速度の比較評価を行ったので報告する.

2. 音高・音源数の同時推定

本研究での音高推定方法は, ハーモニッククラスタリング[7,8]を高速化のために改良した手法を用いている. 即ち, 推定音源数をK個とした場合, まず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から一定以上のパワーを持つピーク位置を複素スペクトル内挿法により抽出する. 抽出されたピークの中で一番低いピークから倍音に対応することが明らかなピークを削除していき, 残ったピークのうち上位 L 個(ただし $L > K$ とする)のピーク周波数 $\{\mu_1, \mu_2, \dots, \mu_L\}$ から選びだした K 個の基本周波数を周波数の低い順に並べて得られるベクトル:

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

に対して, 評価関数:

$$D(\boldsymbol{\mu}) = \sum_{k=1}^K \sum_n \int_{T_n^k} \phi(\omega, n, \mu_k) f(\omega) d\omega \quad (2)$$

の値を求める. ただし $f(\omega)$ は観測スペクトル,

$$T_n^k = \left\{ \omega \mid (n, k) = \arg \min_{m, l} |\omega - m\mu_l| \right\} \quad (3)$$

$$\phi(\omega, n\mu) = (\omega - n\mu)^2 \quad (4)$$

である. $D(\boldsymbol{\mu})$ 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とる. 一方, 基本周波数以外に加え, 削除されずに残っていた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小さいと予想できる. つまり $\boldsymbol{\mu}$ 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 $D(\boldsymbol{\mu})$ の値は減少しを続けるが, 全ての基本周波数が $\boldsymbol{\mu}$ に含まれた後, 倍音や基本周波数以外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のが期待できる. この仮定が成り立つか実際に $D(\boldsymbol{\mu})$ の値の推移を調べてみたところ, 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. $D(\boldsymbol{\mu})$ の変化率 $r(K)$ を

$$r(K) = D(\boldsymbol{\mu}^{K+1}) / D(\boldsymbol{\mu}^K) \quad (5)$$

と定義すると, 変化率が 1 に近い一定値以上の値となった場合, 即ち,

$$r(K) > a : a \text{ は } 1 \text{ に近い定数} \quad (6)$$

を始めて満たした時のKは, 実際の音源数より 1 つ大きい値をとることになる. つまり(3)式を満たした時のKを推定音源数とすれば良い.

3. 結果と考察

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. 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. また補助実験の結果から閾値 $\alpha=0.8$ を用いる.

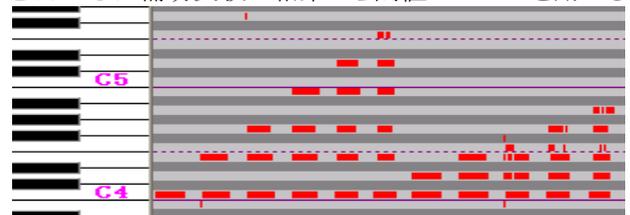


図 1. 1~6 和音混在の認識結果

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

Syou Abe, Hiroyoshi Oda, Toshiaki Matsushima
Dept. of Information Sciences, TOHO Univ.

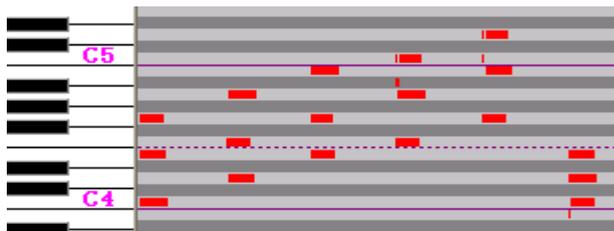


図 2. 3 和音の認識結果

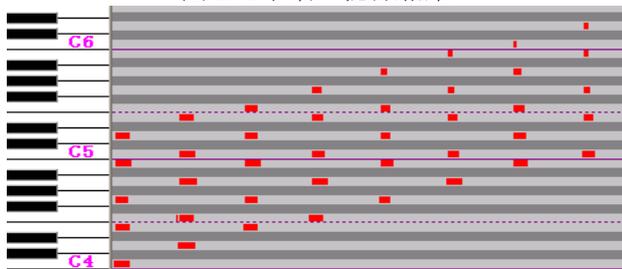


図 3. 5 和音の認識結果

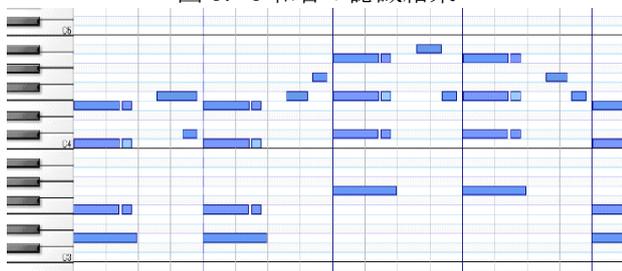


図 4. 正解 MIDI データ

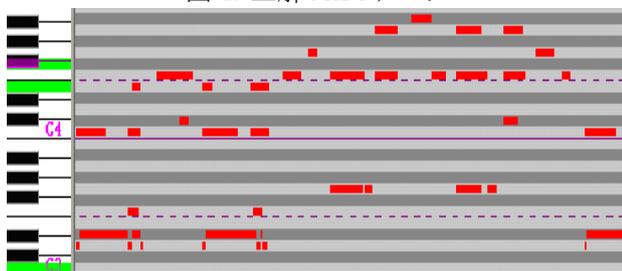


図 5. 採譜結果

まずは認識結果に対するの評価を述べる. 図 1 では 1 から 6 和音まで音源数の推定は正しくできているが, 隣り合った音高のある和音(最後の 3 つの和音)では, ハーモニッククラスタリングにおけるクラスタの範囲が被るため音源数の推定の誤りが増えている. また, 図 3 では高い音高の方が検出できた時間が短くなっていることが分かる. これは高い音高の方がパワーの衰退速度が早いと考えられる.

次に処理速度に対するの評価を述べる. 処理落ちの時間を測るため実験用に作成した単音, 3 和音, 5 和音の音源に対して, 処理に要した時間を測った実験結果を表 1 に示す. 音源の時間に対する処理落ち時間の割合が分かりやすいように発音時間はそれぞれ 10 秒間とした. 単音は C4 の C, C#, D, D#, E, F, F#, G, G#, A, A#, B を, 3 和音は C, C#, D, D#, E, F, F#, G, G#, A, A#, B の各メジャーを, 5 和音に対しては C, C#, D, D#, E, F, F#, G, G#, A, A#, B の各ナインスの音の平均値を記したものである.

	(秒)	単音	3 和音	5 和音
旧アルゴリズム 1(文献[5])	平均	10.0	12.2	15.1
	分散	0.0	0.025	0.162
旧アルゴリズム 2(文献[6])	平均	10.0	10.0	12.2
	分散	0.0	0.0	0.39
新アルゴリズム	平均	10.0	10.0	10.0
	分散	0.0	0.0	0.0

表 1. 処理時間比較

音高が隣り合っている場合や, C5 での音の衰退が激しいための検出漏れが目立った. C2, C3 では音高の検出間違いが多く, 周波数領域が近いためだと考えられる.

図 5 に RWC 研究用音楽データベース[9]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示す. 図 4 は正解データである. 実演奏の音源では音源数の推定の誤りや音高の推定の誤りなどが多く見られる.

4. まとめ

新しいアルゴリズムでは 5 和音までは処理落ちせずに処理することができるようになった. また各和音の処理時間の分散値にもあまりばらつきがみられないため, 和音の組み合わせに対する処理速度の違いはないと考えられるが, ピーク数に処理時間が依存するため, 明らかな倍音と判定できないピークが多い場合などでは処理落ちする可能性がある. 実演奏に対しては誤認識がまだ多く見られるが, 自動譜めくりシステムでは, 楽譜情報が予め分かっているため, 多少の誤認識があっても楽譜とのマッチングを取ることが出来ると予想している. 今後, 自動譜めくりシステムへの組み込みを行ってきたい.

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] 安部翔, 小田良弘, 松島俊明, "自動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", 情報処理学会研究報告書, 2011-MUS92-1(2011)
- [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, pp.27-32(2003)
- [8] 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)
- [9] 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ズ音楽データベース"情報処理学会研究報告, 2002-MUS-44-5, pp.255-32(2002)