社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2008－AVM－61（9）
2008/7/23

ユーザ支援による自動採譜の性能向上

北市 健太郎† 甲藤 二郎†

†早稲田大学基幹理工学研究科情報理工学専攻 〒169-8555 東京都新宿区大久保 3-4-1

E-mail: †{kitaichi, katto}@katto.comm.waseda.ac.jp

あらまし 自動採譜処理において複数音高推定は長年研究されているテーマであるが、現在のところ実用的なシステムはほとんど出来ていない。計算機による自動採譜が難しい理由は、単に計算量だけの問題でなく、計算機の性能向上によって解決されるものではない。そこで、採譜作業を全て計算機に委ねずにユーザと協調して採譜を行う採譜支援システムを提案する。本稿では、ユーザが特定の和音中の最高音を入力することで、その基本周波数以上の周波数帯域をカットする方式を採る。また、コード進行の音楽知識を用いて高精度の推定を実現する。さらにMIDI音源と実音源に対して実際にシステムを構築し、その有効性を確認した。

キーワード 自動採譜，多重音ピッチ推定，コード推定，テンプレートマッチング，ユーザ支援

The performance improvement by user support for automatic transcription

Kentaro KITAICHI† and Jiro KATTO†

† Graduate School of Science and Engineering, Waseda University

3-4-1 Okubo, Shinjyuku-ku, Tokyo, 169-8555 Japan

E-mail: †{kitaichi, katto}@katto.comm.waseda.ac.jp,

Abstract Regarding the automatic transcription, studies on multi-pitch estimation have been pursued for a long time. However, most studies have not implemented a system that can endure practical use at present. It is not solved by the performance gain of the computer because its difficulty is not only the problem of computational complexity. Therefore we propose a transcription-aid system that cooperates with a user without entrusting all. In this paper, the user inputs the highest note in the harmony, and the system cuts the bandwidth over its fundamental frequency. Moreover, higher accuracy is achieved by using the music knowledge of the chord progression. We experimented on estimating the fundamental frequencies of MIDI and actual audio signals and confirmed the effectiveness of these methods.

Keyword automatic transcription, multi-pitch estimation, chord detection, template matching, user support

1. まえがき

コンピュータでオーケストラの生演奏やCD等の音楽データを解析して、その曲を自動的に楽譜化することを自動採譜という。現在市販されている楽譜は、楽曲の作曲家が提供したもの、あるいは音楽的知識を持つ人たちが楽曲を自分の耳で聴いて手動で採譜したものがほとんどである。音楽の経験や知識をある程度持つ人でも、採譜という作業は大変困難である。もし、コンピュータによる自動採譜システムが実現されれば、容易に譜面を作成できるようになる。これにより今まで採譜にかけていた手間が省ける他、作曲支援、自動伴奏など様々なアプリケーションへの応用も期待される。しかし、これまでに様々な手法で研究が成されてきているものの、実用に耐えうるレベルの自動採譜システムは完成に至っていない。

複数音高推定は、自動採譜研究において最もベースとなる研究分野である。柏野らによる、確率モデルに基づくOPTIMA [1]、後藤による、混合音をモデル化し、EMアルゴリズムにより各高調波構造が相対的にどれくらい優勢かを推定する音源数を仮定しない音高推定手法 PreFEst [2]、最も優勢な音高の推定とその高調波成分の除去を繰り返すことで、混合音の構成要素を順次求めていく手法 [3] や、周波数成分を高調波構造の拘束下でクラスタリングする問題と定式化し、各クラスタの重心（音高）と重み（音量）を推定するハーモニック・クラスタリング [4] 等が提案されている。しかし、従来研究では高々3～4程度の音の混合音しか扱うことが出来ず、市販のCDによる音楽の音響信号には有効に機能していないのが現状である。

本稿では、採譜作業全てを計算機に委ねずに人間と協調して採譜を行うシステムを提案する。このようなアプローチを採譜支援システムと呼び、実際にそのシステムを構築し、MIDI音源に対して複数音高推定実験を行い性能の評価を行う。

2. 採譜支援システム

2.1. 概要

従来研究のなされている自動採譜システムは、全ての処理を計算機が行うことを前提としている。しかし、人間が採譜する際に比較的容易に知りえるような情報であっても計算機では抽出困難なものに関しては、計算機で誤った認識を行って精度を下げるよりは、ユーザが計算機を支援してより高精度な認識を得るほうが実用的である。実際、ユーザフィードバックを活用して認識性能を向上させたシステム例は多い。

そこで、計算機とユーザの協調による採譜システムを提案する。提案するシステムの概要を図１に示す。対象となる音響信号は、ユーザの耳によって聴取され、計算機に符号化されたデータとして入力される。採譜を必要とする人は作曲家や演奏者が大半であるため、前提として、ユーザはある程度の音楽経験があるものとする。計算機は楽曲の物理的特徴、すなわち音響信号のスペクトルの特徴を正確に認識することを得意とする。それに対して、ユーザは音楽的な経験や知識を前提とした楽曲の特徴を計算機より的確に捉えられると言えるだろう。計算機とユーザはお互いの得た特徴をやりとりし、最終的に得られたデータを計算機が採譜結果として出力する。

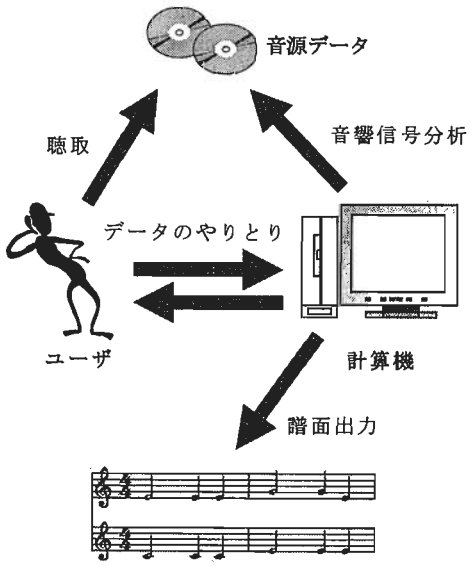

図１：提案する採譜支援システムの概要

2.2. ユーザ支援方法

ここでは、計算機とユーザがどのようなデータをやりとりするかについて検討する。

採譜の精度向上を目的としてはいるが、ユーザによる採譜の作業量が大きくなってしまっては意味がない。本システムでは、

- ユーザの作業量の減少
- 計算機とユーザの効率の良いやりとり

が求められる。ここでは、特に前者について考える。

まず、ユーザのみによる採譜の手順について述べる。対象とする楽曲は弦楽四重奏曲とする。一般的には、拍子・リズム・主旋律の音高・そのハーモニーの音高を順番に採る。個人差はあるが多くの場合、この中で最も時間と労力がかかる作業がハーモニーの採譜である。拍子・リズムは、よほど複雑な楽曲でない限り容易に採譜可能である。また、主旋律は物理的にパワーも大きいことがほとんどで、また、混合音中で最も音程的に高い音である場合が多い。その音高を採ることはそれほど困難な作業ではない。しかし、ハーモニーに関しては同時に複数の音が鳴っており、それぞれのパワーもまばらであることが割合あるので、それぞれの音全てを個別に知覚することは非常に困難である。これは同時発音数が増加するに従って比例的、もしくは指数的に難しくなり、相当の音楽経験や知識がないと採ることができない。

そこで本稿では、音程的に最も高い音（以下、最高音とする）を入力するというユーザ支援方法を採る。前述のように、主旋律の音高を採ることはユーザにとって容易であり、また、主旋律が最高音である頻度は高い。また、主旋律が最高音でなくても、一般的な楽曲で使用される音高の範囲内であれば、人間は聴覚的に低い音より高い音の方が知覚しやすい。以上より、最高音指定という作業はそれほど困難な作業ではなく、ユーザの作業量の減少という要求を満たすものである。

また、計算機側の処理として、音楽知識であるコード進行によって補正する。そうすることで、採譜の精度を向上する。図２に本手法の流れを示す。

2.3. ユーザによる最高音の指定

単音のスペクトルは、その基本周波数においてピークが立ち上がり、さらにその整数倍付近の周波数(以下、倍音周波数)においてもピークが立ち上がる。そのため、単音の場合は周波数構造が予測しやすく、比較的簡単に基本周波数と特定することができる。しかし、多重音のスペクトルでは、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各音源の基本周波数を推定することは難しい。このことは同時発音数が増加するにつれて指数的に増加する問題である。

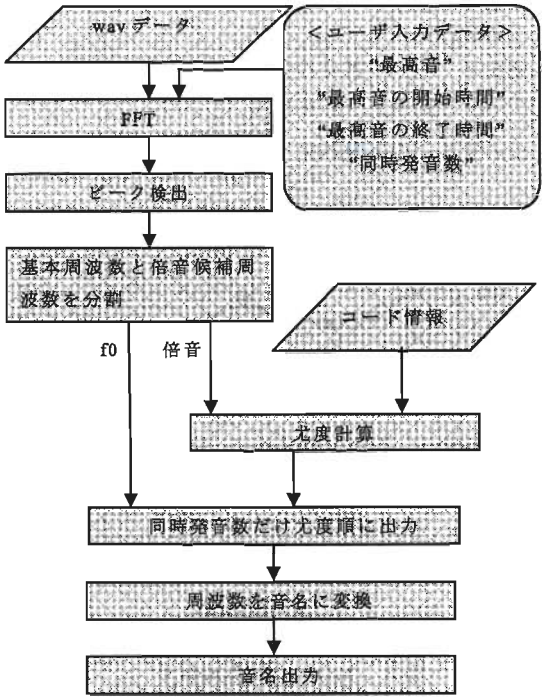

図２：提案手法の流れ

2.3.1. 最高音以上の周波数カット

ユーザが最高音を入力し、その音以上の周波数帯域をカットすることで、残りの周波数スペクトル情報を扱う。これは、単に計算量を削減するためではなく、推定誤りを抑えることが目的である。複数音高推定において、誤検出を起こす原因の多くは２つの問題に起因する。１つ目は、基本周波数のパワーが倍音に比べて小さすぎるため検出できないというもの(missing fundamental 現象)。２つ目は、パワーの高い倍音周波数を基本周波数と認識してしまうものである。最高音以上の周波数をカットすることで、この２つの原因による誤検出を軽減することができる。

2.3.2. 基本周波数と倍音周波数の分割

最高音の基本周波数以下のスペクトルから、複数のピークを検出することができるが、この中には基本周波数と倍音周波数が混在する。ここで、次の仮定を立てる。

（仮定１）倍音周波数は基本周波数の整数倍である

この仮定に基づき、基本周波数と倍音周波数を分割する。あるピーク A の整数倍付近に別のピーク B があった場合、ピーク B を倍音周波数とする。これを周波数の低いピークから順に繰り返し、倍音周波数と認定されなかったピークを基本周波数とする。

このとき問題となるのが、混合音の一部の音の基本周波数のピークが、他の音の倍音周波数のピークと重なってしまうことである。このことは必ずしも起こることではないが、同時発音数や最高音から最低音までの幅の増加に従って発生頻度が増加する。従って、検出された複数の倍音周波数の中から、基本周波数であるものを特定する必要がある。そこで、ピークのパワーと、コード進行による補正により基本周波数を判別する。

2.4. コード進行による補正

2.4.1. コードによる重み付け

まず、各ピークのパワーを指標とすることができるが、基本周波数より倍音周波数の方がピークの大きくなるケースは頻繁に起こるため、必ずしもパワーの大きな点が基本周波数であるとは言えない。そこで次の仮定を立てる。

（仮定２）和音の構成音のパターンはコード進行に密接に関わる

全ての倍音候補の周波数を音名に変換し、コード情報による重みづけを行う。コード推定の詳細については 2.4.2.で述べる。簡略化のため、コードは Major, minor のみを扱う。C Major のときと C minor のときの重み付けを表した表を表１に示す。重み付けの数値は音楽的理論を元に設定した。例えば、コードが C Major のときは、C, E, G の音が最も使用頻度が高く、次に D, F, A, B がよく使用される。オクターブの情報は無視して、これらの 12 種類の音にこの係数を乗算することで、各倍音候補の尤度を求める。

表１：コードによる重み付け対応表

C Major	C	C#	D	D#	E	F
	2	1	1.5	1	2	1.5
	F#	G	G#	A	A#	B
	1	2	1	1.5	1	1.5
C minor	C	C#	D	D#	E	F
	2	1	1.5	2	1	1.5
	F#	G	G#	A	A#	B
	1	2	1.5	1	1.5	1

2.4.2. コード推定手法

楽曲の一部に対して FFT をかけ、そのスペクトル情報から chroma-vector を抽出する。ここでの楽曲の一部は小節単位とし、これについてもユーザが指定する。chroma-vector とは、音高情報のうち、オクターブ違いの同じ音階の成分を全て重ね合わせて 1 オクターブ

内の半音階の 12 音の成分に縮約したものである。楽曲の旋律や和声は，全体をオクターブ単位で上下に平行移動しても調性は変化しないことから，オクターブ方向の分布の情報を取り除く chroma-vector は，調推定に必要な音高情報を圧縮していると言える。

抽出した chroma-vector から、テンプレートマッチングによってコードを推定する。テンプレートには、Gomez[7]が提案したものを使用する。Major と minor 合わせて 24 種類のテンプレートと正規化した chroma-vector の値を比較し、類似度を計算する。24 次元で出力された類似度の中で、最も優勢であるものをその小節のコードとする。

3. 評価実験

まず、評価実験を、MIDI 音源を用いて行なった。MIDI 音源の作成にはシェアウェアの Music Studio Independence[6]を、作成した MIDI ファイルの wav 変換には Timidity[7]を利用した。Timidity を使用することで、一般的に PC に搭載されている MIDI デバイスよりかなり高音質で MIDI を再生・保存（wav, mp3）することができる。

今回の実験で使用した楽曲を以下の表 2 に示す。同時発音数を 3〜7 とし、あらゆる楽器構成での MIDI 音源に対して、本手法の評価実験を行った。なお、実装には MIRToolBox[8][9]を使用した。

表 2 ：使用楽曲

曲名	楽器構成	同時発音数	音源数
Symphony No.9 (Beethoven)	Violin×4	3〜4	32
When you wish upon a star	Alto Sax×2 Tenor Sax×2 Baritone Sax ×1	5	32
Nocturne(Debussy)	Clarinet×3 Oboe×1 Horn×2	6〜7	48

MIDI で作成した楽曲の各発音時間をひとつひとつユーザが指定し、発音ごとに最高音を指定した後、提案手法での複数音高推定を行なった。尚、同時発音数は前提情報として与えている。

また、正解率は以下のように定める。

$$\text{正解率}[\%] = \frac{\text{正解和音数}}{\text{テスト和音数}} \times 100$$

実験は以下の 3 つの手法で行なった。

- 手法 1 ：提案手法（最高音指定とコード推定による重み付け）
- 手法 2 ：コード推定なし（最高音指定のみ）
- 手法 3 ：最高音の指定なし

3 つの手法での実験結果を以下の表 3 に示す。

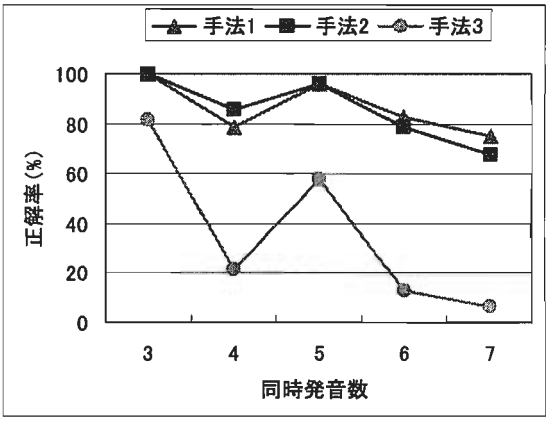

図 3 ：MIDI での実験結果

また、実音源でも同様の実験を行った。使用した楽曲は、現時点では同時発音数 5 の楽曲”When you wish upon a star”のみである。手法 1-3 における正解率を、MIDI と実音源で比較した図を以下の図 4 に示す。

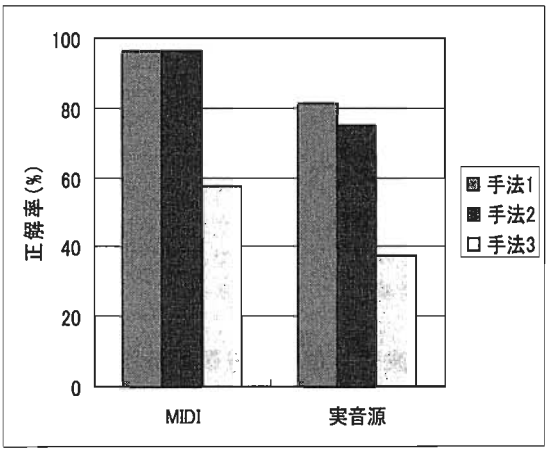

図 4 ：実音源での実験結果

MIDI では音程が正確で一定で正確であるが、実音源ではビブラートなどの奏法などにより音程はある程度揺らぎ、また、音程も正確とは限らない。実際に実音源での実験で出力された音名と正解を比べたところ半音誤りなどが存在した。そのため、正解率は落ちてい

る。しかし、手法1から3までの傾向はMIDIでの実験結果と傾向は同じである。

全体の結果としては、MIDI音源で正解率の平均は手法1で86.46%、手法2で85.6%、手法3で35.85%という結果を得た。また、実音源でもMIDIと同様の傾向が見られる。最高音の入力によってかなりの精度向上を実践することができている。また、コード推定による重み付けにおいても、わずかではあるがより高い正解率を得ることができた。今回の実験では、同時発音数4ではコードなしの方が高い正解率であったが、同時推定音数5,6ではコードありの方が正解率は高くなった。これは、同時発音数が多い方がコードによる重み付けの効果が高くなることを示している。

また、本手法では倍音構造を分析し複数音高推定を行なっているが、統計的手法や学習アルゴリズムは用いていない。複数音高推定の研究でよく用いられているHMM(Hidden Markov Model)や、EMアルゴリズムを用いることで更なる精度向上が期待される。また、コードに関しても、今回はMajor, minor合わせて24種類のコードしか使っていないが、セブンス（C Maj7, C min7など）、sus4、aug、dimなどのコードも考慮に入れることで精度向上を図ることも考えられる。

4. アプリケーション

本手法を用いたアプリケーションのプロトタイプとして、推定結果をピアノロールで表示するソフトウェアを作成した。その画面例を図4に示す。ピアノロールとは、DTMソフトなどで使われる譜面情報の表示方法である。横軸を時間、縦軸を音名としたもので、視覚的に楽曲の流れを表すことができる。本アプリケーションでは、ユーザが入力した最高音を赤、計算機によって推定された音を青で示している。このGUIで編集・修正を行うことによって採譜作業を効率良く行うことができる。

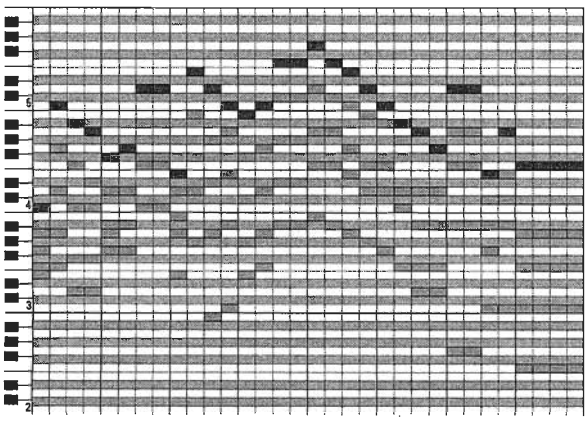

図4：試作アプリケーション画面例

今回の手法では、一つ一つの混合音に対してユーザが最高音をテキストで入力する手法を採ったが、その他の手法も考えられる。楽曲に合わせてユーザがマイクに向かって最高音の旋律を歌うことで最高音の入力を行うことにより、よりユーザによる作業量を減らすこともできるだろう。また、最高音の入力だけでなく、楽曲の繰り返し構造や、使われている楽器の種類などを入力することでより採譜の精度をさらに向上させることも考えられ、応用範囲は広いと思われる。

5. まとめ

本論文では、採譜支援システムの一環として、最高音をユーザが指定することで精度の高い複数音高推定を実装し、様々な同時発音数でのMIDI楽曲・実音源に対してどの程度有効かを確認した。正解率は、同時発音数や楽器の種類だけでなく、音域や和音の校正方法などの楽曲の特性によって大きく左右されるものの、全体的に正解率は向上し、同時発音数の増加による精度低下を軽減することができた。ユーザ支援によって精度向上を実現することが確認された。

文　献

[1] 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, Vol.J79-D-2, No.11, pp.1751-1761. Nov.1996.

[2] 後藤真考, "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, pp.12-22. Jan.2001

[3] A. P. Klapuri, "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness, " IEEE Trans, Speech Audio Process, 11, pp.804-816, Nov.2003.

[4] 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基本周波数抽出における音源数およびオクターブ位置の推定, " 日本音響学会秋季研究発表会講演論文集, 1-1-2, pp.639-640, Sep.2003.

[5] Gomez, "Tonal description of polyphonic audio for music content processing, " INFORMS Journal on Computing, 18-3, pp.294-304, Mar.2006

[6] Music Studio Independence : http://www.frieve.com/musicstd/index.html

[7] Timidity++ : http://sca.uwaterloo.ca/www.cgs.fi/tt/timidity/

[8] Olivier Lartillot, Petri Toiviainen - "A matlab toolbox for musical feature extract , " 10th DAFx(Digital Audio Effects) Conference , pp.237-244. Sep.2007.

[9] Olivier Lartillot, Petri Toiviainen-"MIR in matlab (II): A toolbox for musical feature extraction from audio" ISMIR2007, Sep.2007.

[10] 後藤真孝, 平田圭二, "音楽情報処理の最近の研究," 日本音響学会誌, 60-11, pp.675-681, Nov.2004.

[11] 亀岡弘和, 篠田浩一, 嵯峨山茂樹, "周波数領域の

DP マッチングによる自然楽器演奏の和音ピッチ推定," 情報処理学会研究報告, 2002-MUS46-3, pp.17-22, Jul.2002.

[12]齋藤努, 松井孝誌, 本多英基, 田所嘉昭, “くし型フィルタに基づく DSP を用いたリアルタイム音階検出,” 計測自動制御学会論文集, Vol.34, No.6, pp.504-509, Jun.1998.

[13]谷真宏, 久保田優, 大下隼人, 佃卓磨, 山崎篤史, 北原聡志, 甲藤二郎, "確率推論に基づく自動採譜システムの検討" FIT2004, G-103, Sep.2004.