

## 確率推論に基づく自動採譜システム Automatic Music Transcription based on Probability Reasoning

山崎 篤史† 北原 聡志† 甲藤 二郎†  
Atsushi Yamasaki Satoshi Kitahara Jiro Katto

### 1. まえがき

本論文では、コンピュータを用いて音響信号から自動的に楽譜を生成する、自動採譜について述べる。本研究では、和音の楽曲を入力とし、オンセット検知、音源分離、音高推定、楽器同定の4つの処理から成り、パート判別まで実現する自動採譜システムを提案する。その際、楽器同定処理部については2種の手法の特性比較を行い、最良の方式の組み合わせとしてシステムを構築し、評価を行った。システムは最終的にピアノロール形式の楽譜を出力する。

### 2 自動採譜システムの概要

#### 2.1 システム構成

提案する自動採譜システムの構成を図1に示す。入力にはモノラルの音楽音響信号である。まず、オンセット検知により音の立ち上がり部分を検出し、オンセットごとに入力曲を分割する。次に周波数軸上のスペクトルに対しクラスタリング処理を施し、重複する周波数スペクトルの分離を行うことで、和音から各単音への音源分離を行う。そして、これら各単音について、音高と楽器を推定する。それらに補正処理を施し、統合した結果を楽譜として出力する。

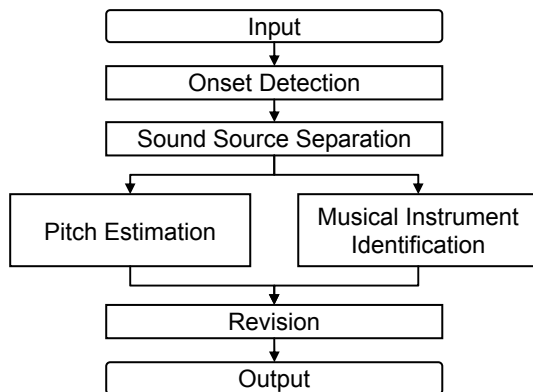


図1 自動採譜システムの構成

#### 2.2 オンセット検知部

本研究では、音楽的知識や音源数、入力音源の周期性等を仮定せずにオンセット一つ一つを認識することができるKlapuriの手法[1]に基づいたオンセット検知を行う[3]。本研究では、手法[1]とは異なる周波数帯域を利用することで、低音域での認識率の改善を図り、さらに、スペクトル形状の時間変化を利用して入力曲を数音符単位に分割する前処理と、スペクトル形状の安定した部分のオンセットを削除する後処理により、音量変化への対応、オンセット過剰検知の抑制を試みた。

#### 2.3 音源分離処理部

音源分離処理部では、亀岡らの「ハーモニック・クラスタリング」[2]に基づいたクラスタリングを用いて音源分離

処理を行う。まず、周波数軸上に広がったスペクトルの形状を混合正規分布で最適近似することを考える。その為に、基本周波数に対応する正規分布の平均 $\mu_k$ だけが自由度をもつ拘束付き混合正規分布によりモデル化し、EMアルゴリズムを適用することで、周波数軸上のスペクトルのクラスタリングを行う。本研究では、EMアルゴリズムの収束を安定・促進させる目的で、周波数軸上のスペクトルに対しピーク抽出を行い、良好な初期値を求める前処理を加えた。

さらに、このクラスタリング処理により得られた情報を利用することで、多重音を各単音成分に分離する。ここで用いる情報は、クラスタリングにより得られた各正規分布の平均（基本周波数） $\mu_k$ と、各周波数成分が各正規分布に属する確率である帰属度 $p_n^k$ である。

$$p_n^k = \frac{w_n^k \cdot g(\omega | n\mu_k, \sigma^2)}{\sum_k \sum_n w_n^k \cdot g(\omega | n\mu_k, \sigma^2)} \quad (1)$$

ここで、 $w$ は各正規分布の重みであり、 $g$ は正規分布を表す。まず、ハーモニック・クラスタリングによって推定された混合正規分布の中の、一つの正規分布と、それに属する周波数成分に注目する。ここで、各階級（周波数）の度数（パワースペクトル密度の値）を度数の合計で割ったものを各階級の相対度数と呼ぶ。ある階級の度数と、その階級の相対度数から、その正規分布の度数の合計が推定される。次に、その合計と各階級の相対度数から、各階級の度数が推定される。これらの処理を混合正規分布全体に施すことで、各単音成分を推定することができる。その後、推定されたこれら各単音成分の比を基に元波形を分割（分配）することで、自然な分離音を得る。

#### 2.4 音高推定処理部

音高推定処理部では、多次元正規分布を用いた音高推定を行う。まず学習作業として、音高ごとの多次元正規分布による楽器音モデルを作成する。次に認識処理として、音源分離処理部で得られた単音にwavelet変換を施し、得られたパワースペクトルを楽器音モデルと比較していき、マハラノビス距離で評価して最も距離の近いものを出力結果とする。

#### 2.5 楽器同定処理部

##### 2.5.1 Support Vector Machine による楽器同定

本研究で扱っている6種類の楽器は減衰系楽器と持続系楽器に分類される。そこでまず、未知の入力音が減衰系と持続系のどちらのカテゴリに属するかSVMを用いて分類を行う。次にそれぞれのカテゴリ内で個別楽器の推定を行う。この個別楽器の推定では、SVMの基本が2値分類器であることを考慮し、楽器ごとに判別用のSVMを用意する。認識対象楽器をクラス1、それ以外の楽器をクラス-1に対応させることで各楽器のモデルを作成し、学習させる。識別では、未知音をすべてのSVMに通し、クラス1と出力されたものを正解楽器とする。2つ以上のSVMで反応があ

† 早稲田大学大学院理工学研究科

った場合は分離面からの距離を比較し、最も大きい値のものを正解とする。なお、特徴量には MFCC とそのフレーム一次差分である  $\Delta$ MFCC、さらにその差分である  $\Delta^2$ MFCC を各 120 次元分ずつ用いる。

表 1 楽器の階層表現

減衰系		cg, eb, pf
持続系	弦楽器	vn
	管楽器	木管
		金管
		tr

## 2.5.2 Dynamic Bayesian Networks を用いた楽器同定

Dynamic Bayesian Networks (DBN) を用いた楽器同定 [8] では、まず、入力音響信号に対して時間 - 周波数変換を行い、楽器音特徴量を抽出する。特徴量は、音楽音響学や楽器物理学の知見等を参考に候補を挙げ、主成分分析により候補を絞り込み決定した。学習処理では、楽器固有の特徴量分布を求め、DBN の知識に相当する条件付確率分布を作成する。ここでは、条件付確率分布として混合正規分布を用いた。認識処理では、入力楽器音の特徴量を観測し、構成したネットワークの子ノード群に証拠として与え、最終的に入力音に含まれる楽器を確率的に推定する。

## 2.6 補正処理部

音高推定処理において誤りやすいものに、オクターブ誤りがある。これは、音響信号が持つ倍音構造により、本来の整数倍の基本周波数を誤認識してしまうものである。本研究では、この誤認識を減少させ補正する処理を加えた。

また、楽器同定処理部では、オンセット毎に分割した単音に対して処理を行っているが、採譜を行うにあたり、「突然メロディ内で楽器の種類が変わることは少ない」という仮定を基に、楽器同定処理の結果にも補正を加えた。

## 3. システム評価

ここでは、図 1 に示す各処理部の評価実験と、それらを統合したシステム全体としての評価実験を別々に行った。学習と実験に使用した楽器音データベースの内訳を表 2 に示す。音高推定処理部及び楽器同定処理部それぞれの評価実験は、学習用と実験用で異なる音源を用いた。また、オンセット検知部及び、自動採譜システム全体の評価実験には、実験用音源を接続することにより作成した最大 2 和音から成る楽曲を用いた。

表 2 使用した楽器音の内訳

楽器	cg	cl	eb	pf	tr	vn	
音域	E2-E5	D3-E5	E1-E3	E1-E5	B3-E5	G3-E5	
音源数	学習用	1296	162	473	441	125	284
	実験用	652	103	224	245	107	142

## 3.1 各処理部の実験結果

各処理部の実験結果を表 3 に示す。オンセット検知部の結果は、式 (3) で、音高推定処理部及び、楽器同定処理部の結果は、単音認識率で評価を行い、式 (4) により求めた。

$$\frac{\text{全オンセット数} - \text{未検出数} - \text{過剰検出数}}{\text{全オンセット数}} \quad (3) \quad \frac{\text{正解音源数}}{\text{実験音源数}} \quad (4)$$

表 3 各処理部の認識率

処理名	認識率
オンセット検知	91.0%
音高推定 (多次元正規分布)	94.9%
楽器同定 (DBN)	86.4%
楽器同定 (SVM)	92.2%

## 3.2 システムの評価実験

各処理部の実験結果を基に、多次元正規分布による音高推定、Dynamic Bayesian Networks による楽器同定を組み合わせてシステムを構築し、評価実験を行った。入力には、器楽の教科書に掲載されている既存の曲を用意した。出力結果の一例を図 2 に示す。

音源同定正解率は、単旋律を入力とした場合 100%、二重奏を入力とした場合 64.4% となった。音源同定正解率とは、音高推定と楽器同定の結果が共に正解した割合である。

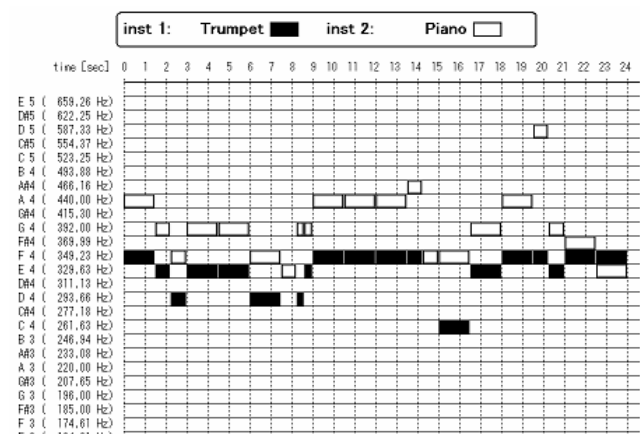


図 2 出力結果の一例

## 4. あとがき

本稿では、オンセット検知部、音源分離処理部、及び、音源同定処理部から成り、楽器のパート判別まで行う自動採譜システムの構築を試みた。

今後は、オンセット検知やクラスタリングに楽器の知識を適用することにより、認識精度の向上を図りたいと考えている。

## 参考文献

- [1] Klapuri et al., "Automatic Transcription of Music", Inproc. Symposium on Stochastic Modeling of Music, 22nd of October, Ghent, Belgium, 2001.
- [2] 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信号音高抽出における音源数とオクターブ位置推定", 情報研報, 2003-MUS-51, pp.29-34, 2003.
- [3] 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.