

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

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†  
M.Tani Y.Kubota H.Ohshita T.Tsukuda A.Yamasaki S.Kitahara J.Katto

### 1. まえがき

近年、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、自動採譜の研究が盛んである。これまで、入力音を和音としても、どの楽器がどのパートを奏でているかまでは判別できないものが多数であった[1]。

そこで、本研究では、和音を入力とし、パート判別まで実現する自動採譜システムの構築を目的とする。システムは主に、オンセット検知部、音源分離処理部、音源同定処理部から成り、確率推論モデルを用いた自動採譜を提案する。

### 2. 自動採譜システムの構成

#### 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。まず、入力音響信号を周波数変換し、周波数成分のクラスタリングを行う。得られたそれぞれのクラスタに関して、周波数帯域別に波形の時間変化を観測し、オンセット(音の立ち上がり点)を検出する。次に、各クラスタに重複して存在する周波数スペクトルの分離を行い、単楽器音を得る。得られた単音から、音高推定で利用する wavelet 変換による周波数スペクトル、楽器同定で利用する種々の楽器音特徴量を抽出する。そして、楽器音をモデリングした多次元正規分布で音高推定、楽器音特徴量をノードとして統合した Bayesian Network で楽器同定を行い、最終的に音源同定結果を得る。

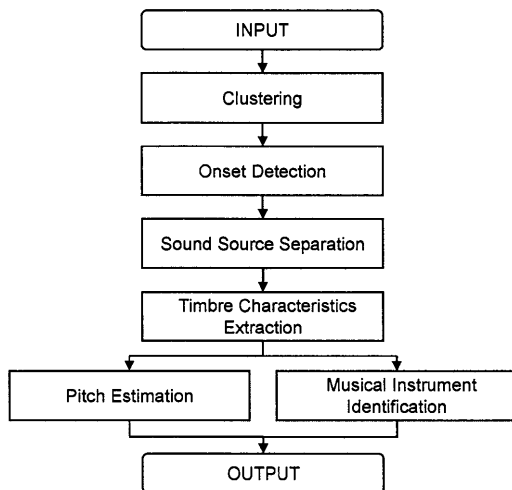


図1 自動採譜システムの構成

Fig.1 The automatic music transcription architecture.

#### 2.2 クラスタリング [2]

クラスタリング部では、亀岡らによるハーモニック・クラスタリング[2]を用い、周波数軸上のスペクトルに対しクラスタリングを行う。これは、時間周波数解析における窓関数や周波数の連続的な変化などの影響によって観測される、周波数軸上に広がったスペクトルの形状を、正規分布で最適近似するものである。これを、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(調波モデル)、EM アルゴリズムにより解く。

また、AIC(赤池情報量規準)[3]を利用して音源数を推定する。まず、予想される発音数より多めの調波モデルを初期配置する。EMアルゴリズムによりモデルパラメータ  $\theta$  の最尤推定値を求めAICを算出し、不必要な調波モデルを削減することを繰り返す。AICが最小となるときの音源数を推定音源数とする。

#### 2.3 オンセット検知 [4]

オンセット検知部での目的は、それぞれのクラスタの正確なオンセット時刻を推定することである。本稿では、音源数や入力音響信号の周期性を仮定せずにオンセット一つ一つを認識する Klapuri のアルゴリズム[4]に基づいてオンセット検知を行った。

このアルゴリズムは臨界帯域に似た 44Hz~22kHz で重ならない 21 帯域のバンドパスフィルタを用いて、周波数帯域別にオンセット要素を検知した後、全帯域の情報を統合して最終的なオンセットを決定する。周波数帯域別のオンセット検知では、まず、バンドパスフィルタの出力を整流・間引きし、100ms のハンギング窓の畳込みにより時間エンベロープ  $A(t)$  を求める。振幅に対する変化量  $W(t) = d/dt(\log A(t))$  の単純なピーク検知により、オンセット要素が推定される。最終的なオンセットは、全帯域の時間的に近いオンセット要素をグループ化し、その音量の閾値処理を行うことで決定される。

#### 2.4 音源分離

音源分離部では、ハーモニック・クラスタリングにより得た平均(基本周波数)  $\mu_k$ 、および、帰属度  $p_n^k$  を利用して、多重音から個々の音源を分離する。

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

ある正規分布に対する帰属度が高い周波数成分に着目し、その成分に対する累積確率とパワー値から、正規分布の母数を推定する。この母数と各成分に対する累積確率から各成分のパワー値を推定する。

† 早稲田大学大学院理工学研究科

‡ 早稲田大学理工学部

### 2.5 音高推定

音高推定部では、学習作業として、wavelet 変換によって得た 120 点のワースペクトルを用いて、平均ベクトル  $\mu$  と共分散行列  $\Sigma$  を求め、120 次元の多次元正規分布で楽器音をモデル化する。

音楽における音階は周波数が対数的に変化する。低周波域で半音の違いを表現するために分解能を上げると、半音での周波数の違いが大きい高周波でも同じ分解能が確保されてしまい、スペクトル表現として効率が悪い。そこで、音高推定部では wavelet 変換を用いる。

認識処理では、音源分離部で得られた単音に対して、wavelet 変換を施しワースペクトル  $\mathbf{x}$  を得る。そして、あらかじめ用意した楽器音モデルの多次元正規分布と比較していき、マハラノビス距離  $D$  で評価して最も距離の近いものを出力結果とする。

$$D^2 = (\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu) \quad (2)$$

### 2.6 楽器同定

楽器同定処理は、統計的性質を持つ楽器音特徴量をノードとして結合した Bayesian Network によって行う。処理の流れを図 2 に示す。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、楽器物理学の知見や、北原らの研究[5]を参考にして候補を挙げ、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、Bayesian Network の知識に相当する、条件付き確率表 (CPT; Conditional Probability Table) を作成する。認識処理では、入力楽器音の特徴量を観測し、構成した Bayesian Network の入力として与え、最終的に入力音に含まれる楽器を確率的に推定する。

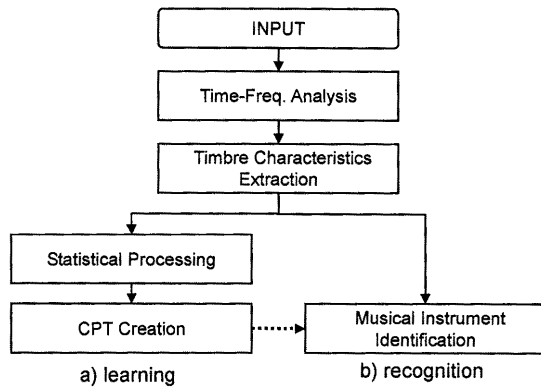


図 2 楽器同定処理部の詳細  
Fig.2 Detail of musical instrument identification.

### 3. 評価実験

ここでは、図 1 に示す各処理部の評価実験と各処理部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[6][7][8]の内訳を表 1 に示す。音高推定部及び、楽器同定処理部それぞれの評価実験は、学習用音源とは異なる実験用音源により行った。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 3 和音から成る楽曲を用いた。

表 1 使用した楽器音データベースの内訳

楽器	eb	pf	tr	cl	cg	vn	
音域	E1-E3	E1-E5	B3-E5	D3-E5	E2-E5	G3-E5	
音源数	学習用	473	385	125	162	1296	284
	実験用	224	245	107	103	651	142

eb : electric bass, pf : pianoforte, tr : trumpet, cl : clarinet  
cg : classic guitar, vn : violin

### 3.1 各処理部の実験結果

各処理部の実験結果を表 2 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、単音認識率であり、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 2 各処理部の認識率

Processing	recognition rate (%)
Onset Detection	76.8
Pitch Estimation	95.7
Musical Instrument Identification	80.8

### 3.2 自動採譜システム全体の実験結果

入力には、器楽の教科書に掲載されている既存の曲を用いた。音源同定正解率は、単旋律を入力とした場合、76.7%、二重奏を入力とした場合、50.0%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。

### 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、楽音全体を捉えた音楽的知識を用いて、システム全体の出力を確率的に補正することなどにより、認識精度の向上を図りたいと考えている。

### 参考文献

- [1] A. Klapuri et al., "Automatic Transcription of Music", Inproc. Symposium on Stochastic Modeling of Music, 22th of October, Ghent, Belgium, 2001.
- [2] 亀岡弘和, 西本卓也, 嵯峨山茂樹, "拘束つき混合正規分布モデルの MAP 推定による同時発話音声の F0 追跡", 日本音響学会, pp.275-276, 2004.
- [3] 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.
- [4] Anssi Klapuri, "Sound Onset Detection by Applying Psychoacoustic Knowledge" IEEE ICASSP, 1999.
- [5] 北原鉄朗, 後藤真孝, 奥乃博, "楽器音を対象とした音源同定: 音高による音色変化を考慮する識別手法の検討", 情処研報, 2002-MUS-46-1, pp.1-8, 2002.
- [6] RWC 研究用音楽データベース,  
<http://staff.aist.go.jp/m.goto/RWC-MDB/index-j.html>.
- [7] Sound Palette,  
[http://forumnet.ircam.fr/rubrique.php3?id\\_rubrique=107](http://forumnet.ircam.fr/rubrique.php3?id_rubrique=107).
- [8] McGill Master Samples,  
<http://www.mcgill.ca/music/events/samples/>.