

Bayesian Network による楽器音特徴量の統合と楽器同定への応用

Integration of Timbre Characteristics by Bayesian Network and Its Application to

Musical Instrument Identification

谷 真宏

甲藤 二郎

Masahiro TANI

Jiro KATTO

早稲田大学理工学部 電子・情報通信学科

School of Science and Engineering, Waseda University

1. まえがき

Bayesian Network は AI 研究の中から生まれた確率推論技術であり、その応用例は、医療診断や遺伝子情報処理、パターン認識（音声認識や文字認識）など様々である[1]。

周波数成分、単音および和音をノードとした Bayesian Network による音源同定手法が報告されている[2]が、本稿では特に統計的性質を持つ種々の音色パラメータをノードとして用い、楽器同定を行う。

2. 従来手法と問題点

現在までに提案されてきた音源同定手法のうちテンプレートを用意する処理では、実音源に対して柔軟性が乏しいことや、処理時間の長さが問題点として挙げられていた。しかし、本提案手法では、楽器特徴量の統計的性質を Bayesian Network のノード間の条件付き確率として保持するので、同一楽器であれば、学習対象外の音源に対しても広く適用でき、また、構成したグラフのノード数に比例する時間で計算できる Belief Propagation アルゴリズムにより処理時間の問題も解決できる。

3. 提案手法

3.1 音色パラメータ

統計処理を行った音色パラメータを表 1 に示す。ここでは試験的に、時間領域に関して 9 種、周波数領域に関して 4 種、ジャンル推定に関して 3 種のパラメータ設定を行った。ジャンル推定とは、入力音が実世界の音楽音響信号である場合、その音楽のジャンルを推定することであり、楽器同定精度の向上を図るために行う。本研究では、ポピュラー音楽、ジャズ音楽、クラシック音楽の 3 ジャンルを対象としている。統計処理には RWC 研究用音楽データベースを利用した[3]。

表 1 統計処理を行った音色パラメータ

時間領域	アタック時間	AT	ディケイ時間	DT	
	サステイン時間	ST	リリース時間	RT	
	最小二乗係数	LS	ピーク比	PR	
	ゼロクロス回数	ZC	包絡線安定度	ES	
	楽器特性				F
周波数領域	倍音数				HO
	基音のx倍以上のスペクトル数				SN
	ピーク比1	F1	ピーク比2	F2	
ジャンル推定	曲中の声の存在				V
	曲の長さ				L
	ジャンル				G

3.2 グラフ構成と動作

提案する楽器同定処理モデルを図 1 に示す。

図 1 は、表 1 で示した楽器の音色に関する特徴量をノード M の子ノード群として結合させたものであり、この子ノード群によりノード M を逐次更新し、最終的に入力音に含まれる楽器を推定する。ただし、ノード M は図 1 に示した 9 種類の楽器を変数としている。

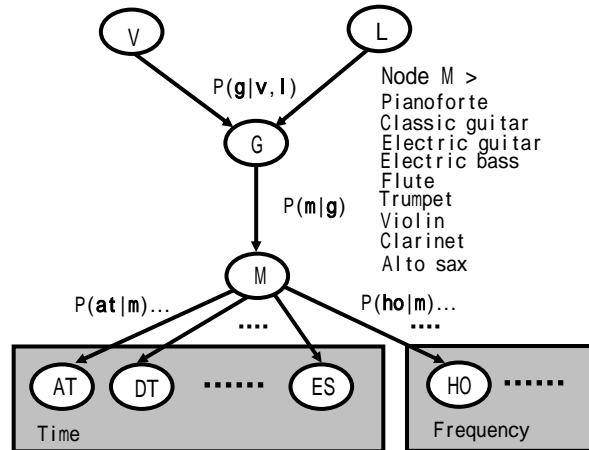


図 1 音源同定処理モデルのグラフ構成

4. 実験

単音楽器同定および 2 和音楽器同定に関してそれぞれ 90 サンプル、36 サンプル用いた。認識結果を表 2 に示す。

入力音の音高範囲は a2(110.00Hz)-c6(1046.50Hz)であり、奏法は主にノーマルで強音、中音、弱音を用意した。

表 2 単音認識および和音認識結果

単音認識率 (%)	96.7	
2和音認識率 (%)	2音認識	33.4
	1音認識	47.2
	誤認識	19.4

5. むすび

本稿では Bayesian Network による楽器同定手法を提案した。

参考文献

[1] Finn V. Jensen, "Bayesian Networks and Decision Graphs," Springer-Verlag, 2001.
 [2] 柏野邦夫他, "音楽情景分析の処理モデル OPTIMA における単音の認識," 信学論, Vol. J79-D- , No.11, pp.1751-1761, 1996.
 [3] RWC 研究用音楽データベース
<http://staff.aist.go.jp/m.goto/RWC-MDB/index-j.html>