F-002

HMMを用いた音源同定アルゴリズムに関する一検討

A Study on Sound Source Identification Algorithm Using Hidden Markov Model

大下 隼人† **甲藤 二郎**†
Hayato Ohshita Jiro Katto

1．まえがき

音情報のみからコンピュータに自動で譜面を生成させる自動採譜というシステムが考えられており、そのシステムを実現するための重要な必要技術の一つとして、音に含まれる楽器の種類と音高を識別する音源同定がある。この音源同定という技術は自動採譜のみならず、コンピュータに音を認識させる全てのシステムにおいて必要不可欠な技術であり、これまで多くの研究がなされているが未だに多数の楽器を正確に識別できるアルゴリズムは見つかっていない。

また一方で、音源同定に類似した音声認識の分野ではHMM(Hidden Markov Model:隠れマルコフモデル)というEM アルゴリズムに基づいた確率統計手法が広く用いられ高い評価を得てきた。

そこで、本稿では周波数解析によって得た楽器音の調波構造の微小時間変動を HMM でモデリングすることでHMM を音源同定に適用し、実音源楽器音での単音認識において提案手法の有効性を検証するための実験を行った。

2．HMM

音声認識の分野で用いられる HMM の形状は、図 1 に示すような left-to-right 型と呼ばれるものが一般的である。中でも離散型 HMM とは、状態 S、遷移確率 a、出力確率 b、の 3 つの要素を持つもので、それぞれ図 1 の S,a,b に対応する。遷移確率 a_{ij} とは状態 S_i から状態 S_j に移る確率のことを言い、出力確率 $b_j(k)$とは状態 S_j においてシンボル k を出力する確率を言う。ここで、出力がシンボルであることが離散型の名前の由来であり、離散型 HMM ではシンボル k がモデルの特性を決める重要なパラメータである。

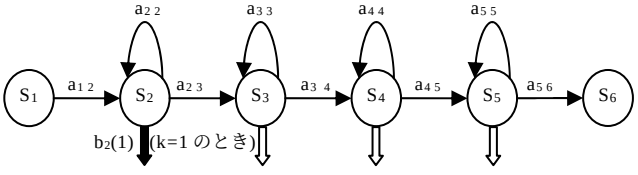

図 1　left-to-right 型 HMM (離散型)

3．GHA

GHA(General Harmonic Analysis:一般調和解析)とは、概周期関数のモデルに基づいた周波数解析手法で、FFT とは違い、解析対象の信号を調和関係にない正弦波を含めた正弦波の集まりとして表現する。またこの特徴により、ある程度高い時間分解能を維持したまま極めて高い周波数分解能が得られるため、非定常信号の解析にも有効であるという優れた手法である[2]。

GHA のアルゴリズムを示すと図 2 のようになる。

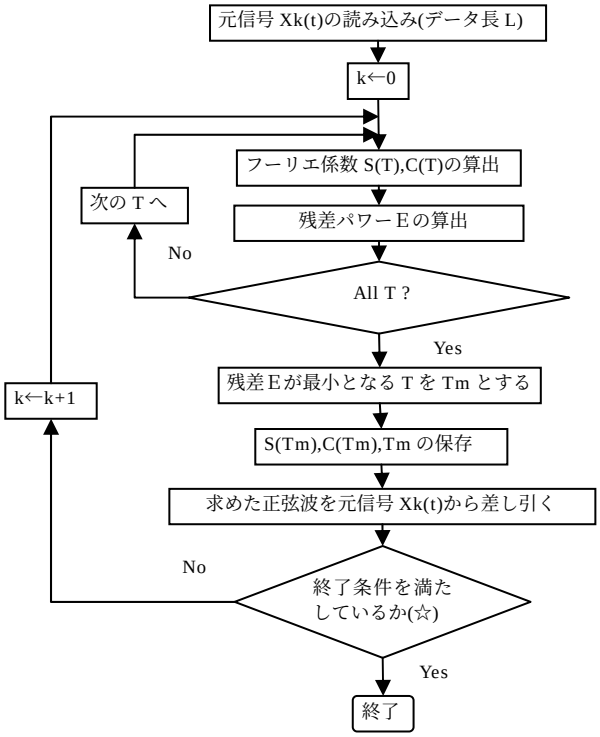

図 2　GHA アルゴリズム

$$S(T)=\frac{2}{nT}\int_0^{nT} x_k(t)\sin(\frac{2\pi t}{T})dt$$

$$C(T)=\frac{2}{nT}\int_0^{nT} x_k(t)\cos(\frac{2\pi t}{T})dt$$

$$e(t,T)=x_k(t)-S(T)\sin(\frac{2\pi t}{T})-C(T)\cos(\frac{2\pi t}{T})$$

$$E=\int_0^L e^2(t,T)dt$$

(☆)終了条件
・必要数の正弦波の抽出の終了
・残差が閾値を下回る
のいづれかを満たす

4．提案手法概要

4.1　全体の流れ

提案手法の全体の流れは、大きく分けて学習処理部(図3)と認識処理部(図 4)から成る。

† 早稲田大学大学院理工学研究科

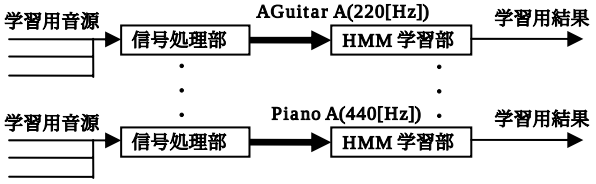

図 3　学習処理部

学習処理部は、ある楽器のある音高に対して一つ用意した HMM を、複数の学習用音源の周波数解析結果の時系列を入力として一つの尤もらしい学習結果を生成するステップである。

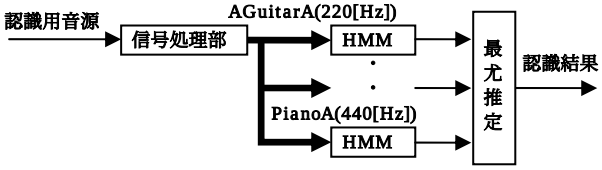

図 4　認識処理部

認識処理部は、ある認識用音源の周波数解析結果の時系列に対し、最尤推定により得られた認識結果が正しいかどうかを判定するステップである。

4.2　信号処理部

信号処理部としては FFT と GHA の 2 タイプを用い、それぞれ別々に実験を行った。

FFT については、一般の楽器音信号が基本音とその倍音成分から構成されていることに着目し、手動で与えた基本周波数に基づいて基本音とその倍音成分のみを取り出した。FFT には 2048 点 FFT (窓長約 45[ms])に Hamming 窓をかけたものを用い、時間変化を少し詳細に観測するために 1024 点(窓長約 23[ms])の overlap を行った。

GHA では周波数分解能を任意に変えることが可能だが、FFT の場合と同様の理由で、A(220[Hz])～G#(6644.88[Hz]) の範囲の音階に相当する周波数のみのパワーを解析し、特徴があるものをシンボルとした。GHA については 1024 点 (窓長約 23[ms])で overlap は無しとして行った。

4.3　HMM

提案手法では図 1 に示した形状の left-to-right 型かつ離散型の HMM を用いた。HMM の個数は、識別したい楽器音の種類(楽器数×音高)の個数だけ HMM を用意し、それぞれ学習用音源により学習した。学習は、初期モデルを学習用音源の統計により構築した後、EM アルゴリズムに基づいた学習法である Baum-Welch Algorithm により行った。その後、学習用音源とは異なった認識用音源を用いて実験を行い、認識用音源に対して演算量が少なく認識に要する時間が短い Viterbi Algorithm により尤度を計算し、最も尤度が高かった HMM に対応する楽器音を認識結果とした。

HMM の要素の中で、具体的な定義をしなければならないのは、モデルの大きさに影響を与える状態 S と出力確率に影響を与えるシンボル k の二つである。

状態 S としては、図 5 に示した楽器音の時間波形のパワー変動に着目して定義した Attack,Decay,Sustain,Release の 4 つと left-to-right 型 HMM を構成するためだけの開始状態、終了状態の 2 つを合わせた 6 状態とした。これにより、状態遷移が時間変動に対応し、音声認識での適用法と同様に楽器音の時間変動を吸収できるモデルとなっている。

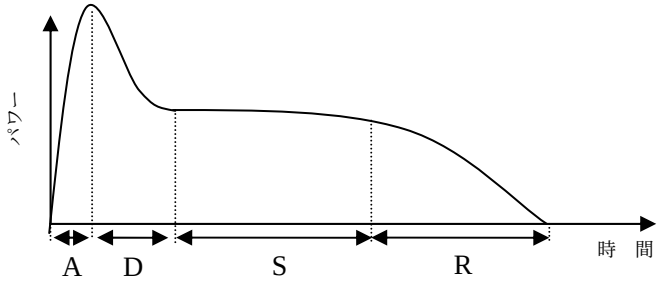

図 5　楽器音の時間波形の概形と状態 (A,D,S,R) の対応

またシンボル k としては、典型的な周波数解析結果(FFT or GHA)を用いた。典型的とは、基本音と倍音成分から成る調波構造の中で特に特徴を持つ形状という意味であり、本研究では、基本音と倍音成分でのパワーの差の総和について閾値処理を行って典型的と判断した。

5．認識実験結果

1 オクターブ A(220[Hz])～A(440[Hz])の 13 音の識別実験を、Acoustic Guitar, Clarinet, Piano の 3 楽器でそれぞれ行った。実験データとしては RWC 研究用音楽データベース[3] 収録のサンプリング周波数 44.1[kHz], 量子化ビット数 16 ビット, モノラルの実音源を用いた。

信号処理部に FFT を用いた場合と GHA を用いた場合の結果をまとめて表 1 に示す。実験結果として示す認識率は、認識用音源のうち、認識用音源の音高と認識結果の音高が正確に合ったもののみ正解として、

$$\text{認識率} = \frac{\text{正解した音高数}}{\text{全ての認識用音源の数}} \times 100\ [\%]$$

によって求めた。

	FFT	GHA
A.Guitar	67.8[%]	76.1[%]
Clarinet	75.0[%]	84.6[%]
Piano	88.5[%]	75.0[%]
総合	72.4[%]	77.6[%]

表 1　単音認識実験結果

6．あとがき

本研究では、音声認識でよく用いられる HMM を音源同定に適用し、楽器音の単音認識において有効であることを示した。また、前処理部である信号処理に GHA を用いる場合の方が、FFT を用いる場合よりも総合で認識率が若干であるが良くなった。今後、音の変動をよりよくモデル化できる連続型 HMM へと発展させる予定である。

参考文献

[1] Lawrence Rabiner, Biing-Hwang Juang, “音声認識の基礎 (下)”, NTT アドバンステクノロジ株式会社, 1995.
[2] 寺田隆彦, “一般調和解析による非定常信号の解析”, PIONEER R&D, Vol.6, No.3, 1996.
[3] RWC 研究用音楽データベース(楽器音データベース) http://staff.aist.go.jp/m.goto/RWC-MDB/index-j.html.