G-29

楽器音の特徴量を利用したフレーズ認識について

Recognition of a short melody using Hidden Markov Model Toolkit

○古谷恵利奈[1]，石原準也[1]
Erina Huruya[1], Junya Ishihara[1]

Abstract: Hidden Markov model toolkit is an integrated package for speech recognition. In this paper, we apply the HTK for the recognition of a short melody played using the guitar and report the experimental results.

１．概要

本論文では，音声認識を対象にしたツールである **HTK** を使用し，ギターで演奏された短いフレーズ認識実験を行い、その結果について報告する．実験では、ギターで演奏した単音を学習データとして用い，その後，「カエルの歌」のフレーズ認識を行った．その結果、do と re の音は認識ができたが，mi と fa の音は，do 又は re に誤認識した結果が得られた．

２．HMM について

Hidden Markov Model（HMM）とは，不確定な時系列のデータをモデル化するための有効な統計的手法であり、主に観測した情報から未知のパラメータを推定するために用いられる．現在では，音声認識、形態素解析（自然言語処理）などに広く応用されている．また，ある入出力系列が与えられた時，どのような状態遷移が行われてきたのかが隠れているため，「隠れ」マルコフモデルと呼ばれる所以ともなっている．その学習には通常 Baum-Welch アルゴリズムが用いられ，状態遷移確率の変化量が閾値以下になるまで繰り返される．一方，認識アルゴリズムには，Viterbi アルゴリズムが用いられ，最も確率が高くなる状態遷移系列だけを求めて近似し，それ以外の計算は打ち切られる．

３．HTK と MFCC

“Hidden Markov Model Toolkit”(HTK)[3] は，**HMM**[1] の構築・学習・認識・評価などといった，パターン認識に必要な工程を一貫して行うためのツールキットである．

HTK は **Table 1** で示されるように主に７つの基本コマンドで構築されている．その使用手順は，まず **HMM** の構成について，状態数・状態遷移数・混合数などを定め，**HMM** 構成ファイルを作成する．次に，**HSLab** コマンドと **HCopy** コマンドを用いて，録音した音声に正解のラベル付けと特徴抽出を行う．そして，**HInit** コマンドと **HRest** コマンドを用いて，**HMM** の初期化と学習を実行する．認識を行うためには，**HParse**，**HVite**，**HResults** コマンドの順にそれぞれ用いる．

Mel Frequency Cepstrum Coefficient (MFCC)とは，一般に音声パターン認識で広く用いられる特徴量である．**MFCC** を求める手順としては，まずフーリエ変換によって求めた音声のスペクトル情報に対し，人間の聴覚特性に合わせたフィルタ群（音声信号処理の引用文献）を用いて対数変換を行う．次に，この対数変換された特徴量の逆フーリエ変換を求め，そうして求められた波形の低周波成分を取り出すことで得られる．

コマンド名	機能
HSLab	音声の録音，ラベル付け
HCopy	特徴抽出
HInit	**HMM** の初期化
HRest	**HMM** の学習
HParse	文法記述をネットワーク表現に変換
HVite	ビタビアルゴリズムによる認識
HResults	認識結果の集計

Table1.　HTK の基本コマンド[2]

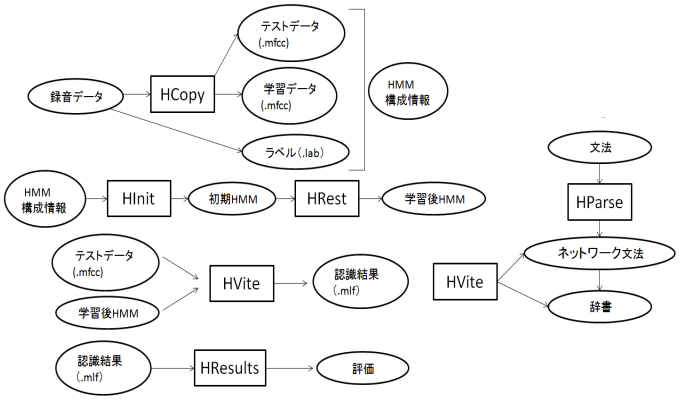

Figure1 HTK を用いた基本的な音声認識システムの構築

１：日大理工・学部・数学

４．フレーズ認識の実験

実験では，**Figure2** に示されるようなギター演奏法に従って発音された各単音をそれぞれ 10 回ずつ演奏し，前章の手順に従い **MFCC** に変換したものを学習データに用いた．

次に，「カエルの歌（ドレミファミレド)」を計 5 回録音し，学習データと同様の手順で得られた MFCC を本研究の目的であるフレーズ認識を行うためのテストデータに用いた．

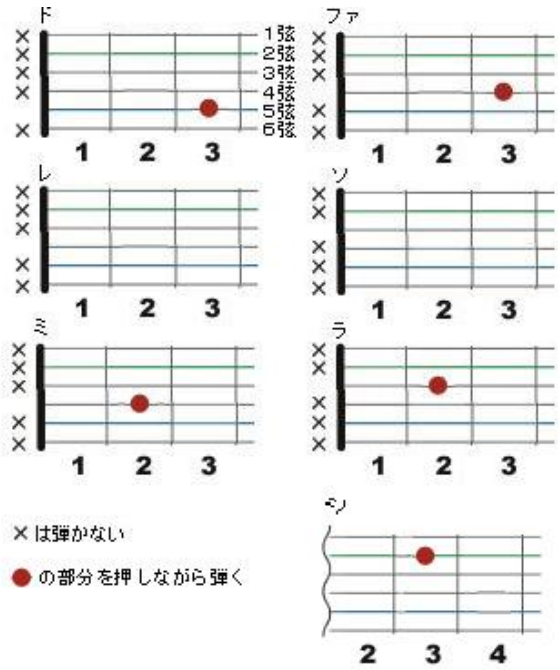

Figure2. 学習データに用いた単音の演奏法

フレーズ認識は，計 7 音で構成される「カエルの歌」のフレーズの一音ごとに波形を手作業で切り抜き，その **MFCC** 特徴データに対して実験を行った．（よって，認識データは，do10 個，re10 個，mi10 個，fa5 個の計 35 個のデータである．）

なお, 実験では, 演奏データ録音の際にギター（ARIA PRO Ⅱ）を用いた．また録音機材としてボイスレコーダー（OLYMPUS 製 VoiceTrekDS-750）を使用し，サンプリング周波数 44.1kHz、量子化ビット数 16bit にて録音した．

５．実験結果

Table 2 および Table 3 に実験結果を示されるように，do は 100%認識されたのに対し，re は 80%の認識率であった．その一方，mi は 8 個が do, 2 個が re と誤認識し，fa については，1 個が do, 4 個が re と誤認識した．つまり，mi と fa 共に全て誤認識されてしまった．

入力＼出力	do	re	mi	fa
do	100	0	0	0
re	20	80	0	0
mi	80	20	0	0
fa	20	80	0	0

Table 2. フレーズ認識の混同行列（単位 %）

６．実験結果の考察および結論

本研究では，**HTK** を用いて「カエルの歌」のフレーズ認識実験を行った．実験では、do のみが全て正しく認識されるという結果が得られた．一方，mi と fa 共に全て誤認識されてしまった．誤認識した一番の原因としては，学習データの数が不十分だったと考えられる．そのため，学術講演会当日では，学習データ，評価用データを共に数を増やし，その結果について報告する予定である．

また，「カエルの歌」だけでなく，他にもいくつか複数のフレーズについても認識実験を行い、その結果について報告する予定である．

７．参考文献

[1] 荒木雅弘：「フリーソフトでつくる音声認識システム」，森北出版株式会社，pp.22-23 , 130-147(2007).

[2] 中田圭介：「Hidden Markov Toolkit を用いた単語音声認識について」，日本大学理工学部学術講演会論文集 (2010).

[3] Hidden Markov Model Toolkit のダウンロード先：HTK Speech Recognition Toolkit(http://htk.eng.cam.ac.uk/)