

E-076

## 音声情報を用いた夜行性鳥類の種識別

岩崎 祐介† 三田 長久† 牧野 洋平† 高橋 幸司† カムケオ シーパチャン†  
Yusuke Iwasaki Nagahisa Mita Yohei Makino Koji Takahashi Khamkeo Sypachanh

## 1. 研究背景

近年、急激な技術革新に伴い、環境問題が深刻になっている。それにつれて、環境評価の研究が盛んに行われている。本研究では、環境の変化に敏感な野鳥について、鳴き声を使って種識別を行う方法を検討した。対象を夜行性野鳥とした場合の鳴き声による種識別の結果について報告する。

## 2. メルケプストラム分析

本論文では、通常の音声認識に用いられるメルケプストラム(MFCC)係数を利用して識別を行った。FFTによるスペクトルを元に、メルスケールの帯域フィルタ群出力を抽出する手順を図1に示す。この処理によって得られる係数をMFCC係数という[1]。また、取り出された係数の個数をMFCCの次数と呼ぶ。

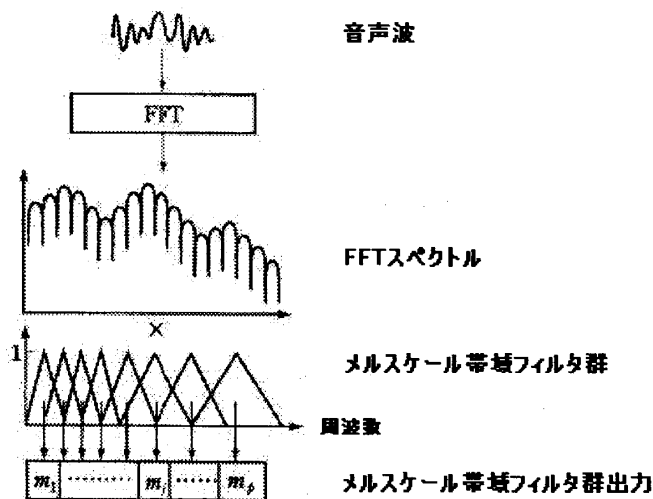


図1 MFCC係数の抽出手順

メルスケールは、式(1)で定義される。

$$Mel(f) = 2595 \log_{10} \left( 1 + \frac{f}{700} \right) \quad (1)$$

いま、各帯域フィルタの出力を  $m_j$  とする。このとき、MFCC係数は、離散コサイン変換(DCT)を用いて式(2)で計算される。

$$c_i = \sqrt{\frac{2}{N}} \sum_{j=1}^N m_j \cos \left( \frac{\pi \cdot i}{N} (j - 0.5) \right) \quad (2)$$

スペクトル情報を表すパラメータの、 $n$ フレームにおける  $i$ 番目の値を  $c_i(n)$  と表す。このとき、時刻  $n$  を中心とした区間  $[n - \delta, n + \delta]$  における  $c_i(n)$  の値に、直線を当てはめた場合の直線の傾きを  $\Delta c_i(n)$  と表すと、式(3)が成り立つ。

$$\Delta c_i(n) = \frac{\sum_{k=-\delta}^{\delta} k \cdot c_i(n+k)}{\sum_{k=-\delta}^{\delta} k^2} \quad (3)$$

$\Delta c_i(n)$  は、 $c_i(n)$  の時間的な変化量(動的特徴)を表すものである。この  $\Delta c_i(n)$  を  $\Delta$ MFCC係数と呼ぶ。

## 3. 特徴抽出について

種識別のための特徴抽出を次に示す手順で行った。

- (1) 野鳥の音声は1秒間の音声データを用いた。これは、1秒間あれば野鳥の識別に十分なデータが得られるためである。
- (2) 野鳥の音声にフレームを半分ずつ重ならせて、短時間スペクトル分析を行う。
- (3) スペクトルサブトラクション法で雑音の除去を行う。
- (4) 切り出したフレームごとにメルスケール帯域フィルタ群を乗算する。帯域数は16とした。
- (5) 式(2)よりMFCC係数を求める。
- (6) MFCC係数から式(3)を用いて  $\Delta$ MFCC係数を求める。
- (7) フレームごとにMFCC係数が16個出ているので1秒間の全フレームの同じ帯域部分の平均を求める(手法1)。平均をとる際、絶対値をとって平均する場合を手

† 熊本大学大学院自然科学研究科情報電気電子工学専攻

法2とする。

- (8) 16個のMFCC係数が求まる。
- (9) フレームごとに $\Delta$ MFCC係数が16個出ているので1秒間の全フレームの同じ帯域部分の平均を求め、平均を求めるときには、 $\Delta$ MFCC係数の絶対値をとり、全フレーム数で割る。
- (10) 16個の $\Delta$ MFCC係数が求まる。
- (10) MFCC係数、 $\Delta$ MFCC係数をそれぞれ最小値0、最大値1として正規化する。
- (11) 合計32個の特徴が求まる。

#### 4. ニューラルネットワーク

識別にはニューラルネットワークを用いた。今回は、3層構造とし、教師あり学習である誤差逆伝播法を用いた。入力ユニット数32、中間ユニット数32、出力ユニット数13として識別を行った。図2に3層ニューラルネットワークの概略図を示す。

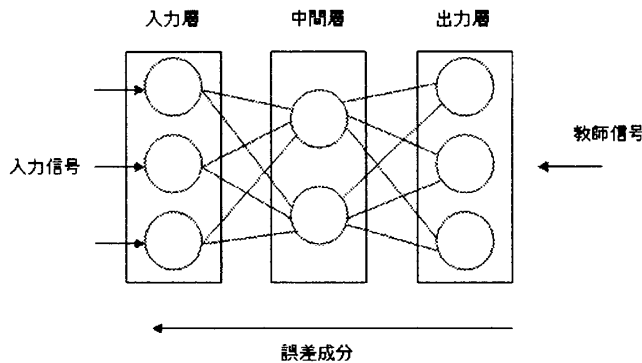


図2 3層ニューラルネットワーク

#### 5. 識別結果

野鳥の種類は、森林性の夜行性鳥類である、アオバズク、アカハラ、オオコノハズク、コノハズク、トラツグミ、トラフズク、フクロウ、ホトトギス、マミジロ、ミゾゴイ、ヤマシギ、ヨタカ、ジュウイチの13種で識別を行った。図3に、手法1と手法2を用いた識別結果を示す。

ニューラルネットワークの教師数は110、識別したサンプル数は354である。

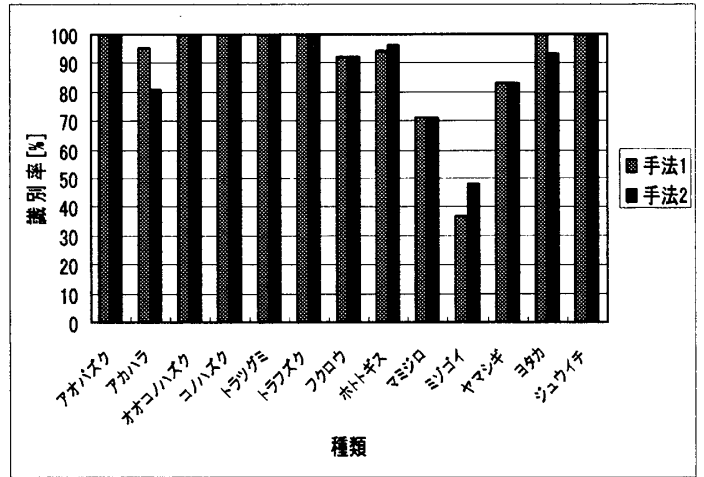


図3 手法1と手法2の識別率の比較

手法1, 手法2ともに全体の識別率は90%であり、識別率の差は見られなかった。

#### 6. まとめ

結果としては、識別率100%の野鳥が複数あり、全体で識別率90%と良い結果となった。しかし、ミゾゴイの識別率が悪く、鳴き声の似ているトラフズクと誤識別することが多くあり、まだ改善する必要がある。

本研究は一部を環境技術開発等推進費の補助を受けて実施しました。

#### 参考文献

- [1] 安藤彰男, “リアルタイム音声認識”, 電子情報通信学会, 2003