E-078

野鳥の音声データの圧縮による種識別への影響の検討

Examination of influence on species identification by compression of wild bird's voice data

高橋　幸司‡　三田　長久‡　牧野　洋平‡　岩崎　祐介‡　カムケオ・シーパチャン‡
Koji Takahashi　Nagahisa Mita　Yohei Makino　Yusuke Iwasaki　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．現在、環境の変化に敏感な野鳥、特に夜行性野鳥の鳴き声を録音、自動識別し、分布状況を調査するシステムを構築している。それに伴い、長時間の音声データを録音する必要があるが、長時間の音声データとなるとデータ量が大きすぎるため、録音時の音声データの圧縮が必要となる。本論文では、夜行性野鳥の音声データを MP3 および ATRAC を用いて圧縮、それを用いて種識別し、圧縮せずに種識別した場合の識別率を低下させること無く、どこまで圧縮できるかということを検討した。

なお、識別には東谷等の方法[1]を用いており、識別対象種は、森林性の夜行性野鳥である、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 種である。

2．音声圧縮

MP3[2]では、人間の聴覚特性を利用して、人間が知覚し辛い音を取り除くことでその圧縮率を高めている。つまりマスキング効果によって聴き取れなくなる音の情報量を捨て去ることで圧縮率を高めている。この他には、人間には可聴領域というものがあり、可聴領域外の音のデータも捨て去りことで、さらにその圧縮率を高めている。

ATRAC[3]とは、ソニー株式会社が開発した MD のための音声圧縮方式で、ATRAC3 や ATRAC3plus 等がある。ATRAC3 では 0~22 kHz の周波数帯域を均等に 4 分割し、それぞれを MDCT にかけ時間-周波数変換を行いそれらをソニー独自の符号化アルゴリズムで符号化し圧縮している。ATRAC3plus では 0~22 kHz の周波数帯域を MP3 同様、32 の帯域に分割し分解能を高めており、符号化の部分でも ATRAC3 よりもさらに複雑な符号化アルゴリズムを用いている。本論文ではこの ATRAC3plus を用いている。

3．識別方法[1]

識別にはニューラルネットワークを用いており、今回は、3 層のニューラルネットワークを利用し，誤差逆伝播法で学習を行った。入力ユニットは 32、出力ユニットは野鳥の種類である 13 種類で行っている。中間ユニットは、入力ユニットの 32 にあわせて、今回識別している。

まず、1 秒間の鳴き声を準備する。そして、周波数をある特定の帯域ごとに分割して、帯域パワーを求める。ここで重要になるのが周波数の分割の方法である、夜行性の野鳥は低い周波数領域で鳴く野鳥が多いため、低い周波数の分割数を大きく、高い周波数の分割数を少なくした。決定した分割を示す。

0.25～1[kHz]	0.25[kHz]間隔
1～3[kHz]	0.4[kHz]間隔
3～5[kHz]	0.5[kHz]間隔
5～8[kHz]は,	0.75[kHz]間隔

また、野鳥の声の高さの個体差に対応するため、図 1 による窓関数をかけて計算した。

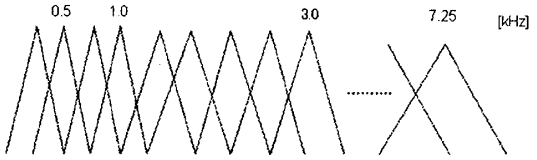

図 1　帯域パワー導出時に使用した窓関数の図

短時間フーリエ変換(2048pt)ごとに、周波数帯域パワーの値を求め、平均した値と微分値をニューラルネットワークの入力とした。平均は最大値 1 になるように正規化を行い、微分合計値は式(1)で計算を行った。これは、サンプリング周波数の変化に対応できるようにしており、単位時間当りの結果になっている。

$$\text{微分合計値} = \frac{\text{微分値の絶対値の合計}}{\left(\frac{\text{ウインドウのサンプル数}}{\text{サンプリング周波数}}\right) \times \text{微分の計算回数}} \quad (1)$$

4．シミュレーション

シミュレーションの流れを図 2 に示す。

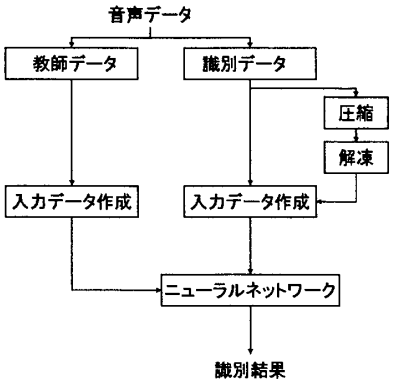

図 2　シミュレーションの流れ

‡熊本大学大学院 自然科学研究科 情報電気電子工学専攻

図 2 に示すように、一度圧縮・解凍したデータと圧縮せずそのままのデータそれぞれから、周波数帯域パワーの平均および微分合計値を計算し、ニューラルネットワークへの入力データとした。この入力データの段階での圧縮と非圧縮データの違いをその比率で図 3、図 4 に表す。図 3 は MP3 を用いており、図 4 は ATRAC3plus を用いている。今回はアオバズクを例に挙げている。

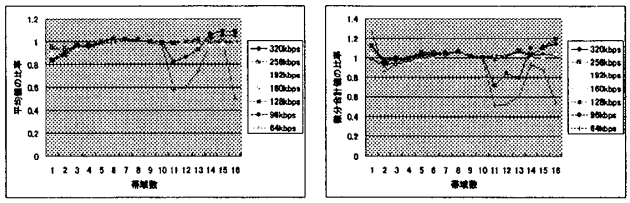

図 3　平均値・微分値の比率(MP3)

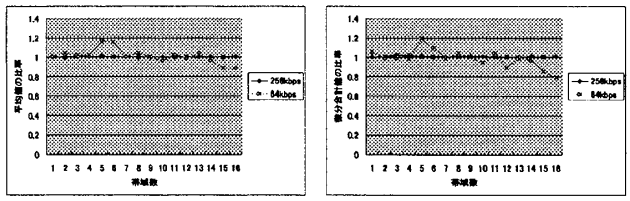

図 4　平均値・微分値の比率(ATRAC3plus)

次にシミュレーション条件について述べる。今回は、13 種類の野鳥の識別を行った。ニューラルネットワークのパラメータを以下に示す。

学習データ	110 個 (1 種につき 6～10 個)
中間ユニット	32
学習係数	0.05
学習回数	50000

5. 結果

前章のシミュレーション条件による結果を以下に示す。

ビットレート(kbps)	識別率(%)
元データ	93.8
MP3	
320	93.3
256	93.5
192	93.5
160	93.3
128	93.5
96	93.3
64	93.3
ATRAC3plus	
256	93.8
64	94.1

MP3、ATRAC3plus の各ビットレートで識別したところその識別率は、元データで識別したときの識別率と比べて、差は-0.5~+0.3 %となり、ほとんど元データで識別したときと変わらない結果となった。ATRAC3plus の 64 kbps で圧縮・展開したデータでの識別率は元データの識別率を上回る結果となった。これは、圧縮の際の量子化の部分で、鳴き声と同時または、その前後に存在する低い雑音が消去されたことなどが理由として考えられる。

また、各ビットレートで圧縮したデータのデータ量を以下に示す。

各データ	データ量(kB)	圧縮率(%)
元データ	172	0
MP3		
320	40.8	76.3
256	32.6	81
192	24.5	85.8
160	20.4	88.1
128	16.3	90.5
96	12.2	92.9
64	8.2	95.2
ATRAC3plus		
256	37.9	78
64	11.9	93.1

一晩中(8 時間とする)1 GB のディスクに録音する場合、入れることが出来る 1 秒データの最大データ量は 34.7 kB となるので、表 4.14 のうち可能なものは、MP3 では 256 kbps 以下、ATRAC3plus では 64 kbps となる。どのビットレートでも、かなり元データの識別率に近い識別率を示していることから、識別用のデータを録音する際に上記のビットレートで圧縮しても十分識別可能であることがわかる。

6. まとめ

本論文では、MP3、ATRAC3plus という現在最も普及している圧縮形式を用いて夜行性野鳥の 1 秒間の音声データを圧縮・展開し、圧縮していないデータで識別した場合の識別率にどのような影響を与えるのか検討した。結果としてはどのビットレートで圧縮しても、識別率にはほとんど影響を及ぼさないことが判った。また本実験により MP3 では 256 kbps 以下、ATRAC3plus では 64 kbps 程度のビットレートであれば、一晩中(8 時間とする)録音することが可能で、そのデータを用いて識別しても十分な識別結果が得られることが判った。

今後の課題としては、今回は 13 種の野鳥で識別・検討を行なったが、さらに多くの種を識別するための、野鳥の鳴き声の特徴の抽出や識別手法の改善、またそれらの種のデータを圧縮・展開した場合の検討などが挙げられる。

なお、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た。

参考文献

[1] 東谷幸治　三田長久　牧野洋平　「音声情報を用いたニューラルネットワークによる野鳥の種識別」　2007 年電子情報通信学会総合大会　D-14-10

[2] 守谷 健弘　「音声符号化」　コロナ社　2000 年

[3] http://www.sony.co.jp/Products/ATRAC3/　SONY ATRAC情報サイト