

# 奄美大島に生息する希少種の鳴き声 自動認識に関する研究

眞島 京音<sup>1</sup> 福元 伸也<sup>1</sup> 鹿嶋 雅之<sup>1</sup> 渡邊 睦<sup>1</sup>  
鵜川 信<sup>2</sup> 榮村 奈緒子<sup>2</sup>

**概要**：2021年に奄美大島が世界遺産に登録されたことに伴い、生息する野鳥の観測が必要となった。現在行っている観測方法は手動で重労働である。そこで、鳴き声から野鳥の存在を確認できるシステムの開発を行うこととした。本研究では、野鳥の音色特徴を表しているメルケプストラム（MFCC）係数と声紋を通して鳴き声を観測できるスペクトログラム画像を特徴に用いて、全9種の識別を目標にする。その際、鳴き声の近い鳥についても分類できるようにする。結果として、6種のみではあるが、高い精度で分類することができた。今後は残りの野鳥についても対象とし、さらなる精度向上にむけてシステムの構築を目指したい。

**キーワード**：機械学習、ニューラルネット

## Research on automatic recognition of calls of rare species inhabiting Amami-Oshima Island

KYOTO MASHIMA<sup>1</sup> SHINYA FUKUMOTO<sup>1</sup>  
MASAYUKI KASHIMA<sup>1</sup> MUTSUMI WATANABE<sup>1</sup>  
SHIN UGAWA<sup>2</sup> NAOKO EIMURA<sup>2</sup>

**Abstract**: With the registration of Amami Oshima Island as a World Heritage Site in 2021, it has become necessary to observe the wild birds that inhabit the island. The current observation method is manual and hard work. Therefore, we decided to develop a system that can confirm the presence of wild birds based on their calls. The goal of this study is to identify all nine species using the Melkepstrum (MFCC) coefficient, which represents the timbre characteristics of wild birds, and spectrogram images, which allow the observation of calls through voiceprints, as features. In doing so, we will also be able to classify birds with close vocalizations. As a result, we were able to classify only six species with high accuracy. In the future, we would like to target the remaining bird species, and aim to build a system to further improve the accuracy.

**Keywords**: Machine Learning, Neural Networks

### 1. 研究の背景と目的

2021年に奄美大島は「生物多様性」を理由に世界遺産に

登録された。これは固有種や希少生物の存在に高く評価されているからである。登録を維持するためには、生息する野鳥の観測を行い、希少種の存在を確認し続けなければならないという背景がある。

現在、大量の音声データを人の耳で聞き分けて、鳥の存在を確認している。本研究ではこのような作業を軽減し、

1 鹿児島大学 工学部  
Faculty of Engineering, Kagoshima University, Japan  
2 鹿児島大学 農学部  
Faculty of Agriculture, Kagoshima University, Japan

自動で野鳥を識別できるシステムの開発を行うことを目的とする。

## 2. 外部状況

関連研究では、メルケプストラム (MFCC) 係数を特徴量に用いた野鳥 13 種の識別を行っている。識別率は全体で 90~100%と高い精度が出ているが、鳴き声が似ている鳥については 40~50%と低い精度となっている[1]。

本研究では、鳴き声の似ている鳥についても分類することを目的とし、時間的变化や周波数特性を見ることのできるスペクトログラム画像 (図 1) の特徴を加味してより優れた識別を行う。

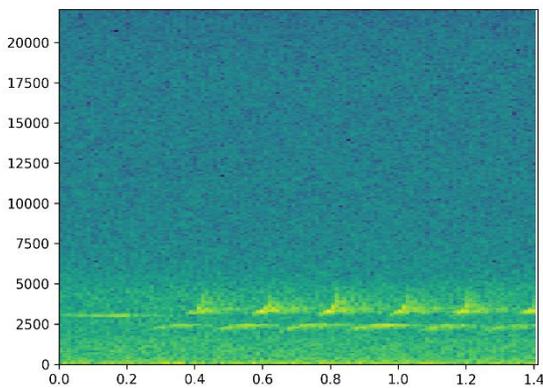


図 1 アカヒゲのスペクトログラム画像  
Figure 1 Spectrogram image of Redbeard.

高橋が行った「津軽弁の音声認識」という研究[10]があった。津軽弁には、母音の無声化、カ行タ行の有声化 (濁る)、シとスの区別がつかないなどの特徴があるようだ。先行研究では、方言という音声特性をスペクトログラム画像によってあらわし、それを機械学習によって、画像認識を行っている。結果として、「K-means によるクラスタリングでかなりの精度で画像を認識できた」とある。

先行研究[3]より、奄美大島のオオトラツグミについて、スペクトログラム画像にて存在を確認しており、本画像を用いての識別方法は無難なものだと考える。当研究より、他の鳥や動物における鳴き声の存在確認にはスペクトログラム画像は有効であるといえる。

本研究でも、鳥の鳴き声についてスペクトログラム画像にて特徴をもとに画像識別を行う。

先行研究[1]では、13 種の鳥について、1 秒で区切るという方法を用いていた。しかし、鳥の鳴き声には、ウグイスの「ホーホケキョ」のように、代表される鳴き方というものがある。先行研究[8]では、ヒヨドリについては「ヒーヨ」という区切り方をし、キジバトについては「ホーホーホーホー」という区切り方をしていた。本研究でも使

用する鳥の鳴き声のデータについて、そのような鳥固有の鳴き方の区切りを意識し、音声データを使用する。

## 3. 実験方法等

### 3.1 データの採取

本研究は、鹿児島大学農学部との共同研究である。データについては奄美大島の現地より、鹿児島大学農学部の柴村先生が録音したデータをもとに、研究を進めた。

録音時間は朝昼の 3 時間の録音である。

場所については役勝唐川、役勝赤畑、役勝嘉徳又・長田、役勝丸畑、はつの、三太郎、金作原、嘉徳川内山、網野子亥ノ川である。役勝嘉徳又・長田、嘉徳川内山、網野子亥ノ川以外の箇所はそれぞれ設置場所の地形を尾根と谷に分けた。

さらに標高は 50~300m で観測している。

録音対象は、奄美大島の固有種についてアカヒゲ・オオトラツグミ・オーストンオオアカゲラ・ルリカケス・カラスバト・リュウキュウコノハズク・ズアカアオバト・リュウキュウキジバト・アオハズクである。そのほかにもキツツキを録音している。オーストンオオアカゲラとキツツキに関しては鳴き声以外にドラミングについて録音されている。

本研究では、アカヒゲ・オオトラツグミ・ルリカケス・カラスバト・リュウキュウキジバト・ズアカアオバトの 6 種について識別を行うこととする。オーストンオオアカゲラについてはデータ数が 15 個しかなく、データ不足のためである。リュウキュウコノハズクとアオハズクについては鳴き声がおよそ 0.1 秒と短すぎるため、现阶段で区切る点をみつけきれないためである。

### 3.2 提案手法

本研究は以下 (図 2) の構成に沿って識別を行う。

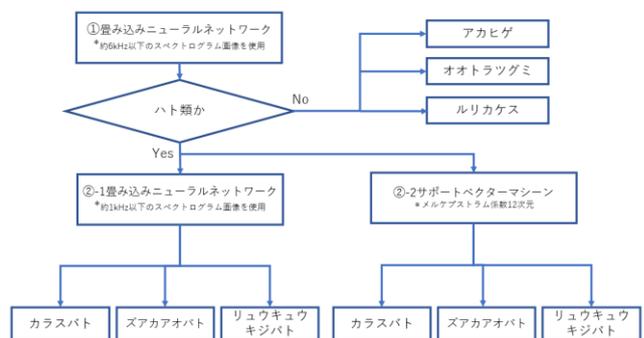


図 2 野鳥識別モデル

Figure 2 Bird Identification Model.

①野鳥の音声から生成したスペクトログラム画像を使用し、畳み込みニューラルネットワーク (CNN) で学習したデータをもとに画像で識別を行う。ここでは多くの野鳥の周波数帯域である 0~6kHz あたりを対象にし、特徴が大きく

でる3匹の鳥とハト類の分類を行う。つまり6種の鳥を4種へと分類する。

使用する画像は以下のようなものである(図3)。

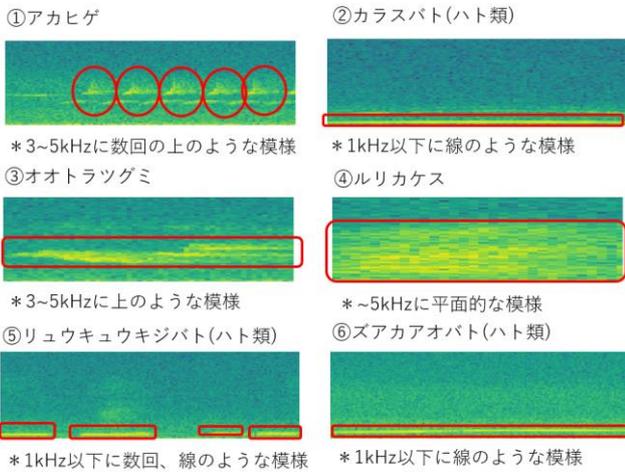


図3 実験①で使用するデータ

Figure 3 Data to be used in experiment ①.

アカヒゲ、ルリカケス、オオトラツグミについてはどの画像も似たものであり、特徴がわかりやすくでている。この中で、区別が付きにくいのはカラスバト、リュウキュウキジバト、ズアカアオバトである。

ハト類についてはさらに②-1や②-2で詳細に分類する。

②-1ではハト類に特化した周波数帯域の画像にて、①同様の処理を行う。

使用する画像は以下のようなものである(図4)。

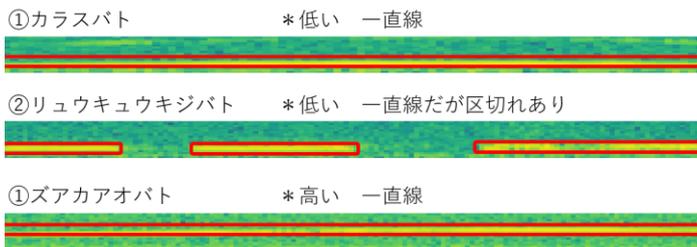


図4 実験②-1で使用するデータ

Figure 4 Data to be used in experiment ②-1.

②-2では音色を特徴量とするメルケプストラム(MFCC)係数12次元の値をもとにSVM(サポートベクターマシン)にて識別を行う。

これは、スペクトル概形をフィルタにかけて圧縮して算出した値である。図5,6,7はハト類各種について50データのメルケプストラム(MFCC)係数を表したものである。カラスバトとリュウキュウキジバトについてはおよそ右下がりの概形、ズアカアオバトについてはU字のようなカーブを描く概形である。上記特徴を用いて、実験を行う。

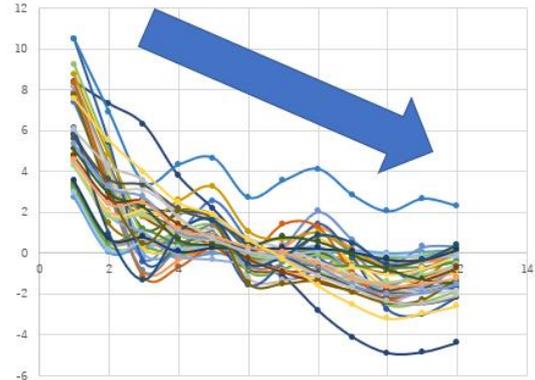


図5 カラスバトのMFCC係数

Figure 5 MFCC coefficient for crow pigeons.

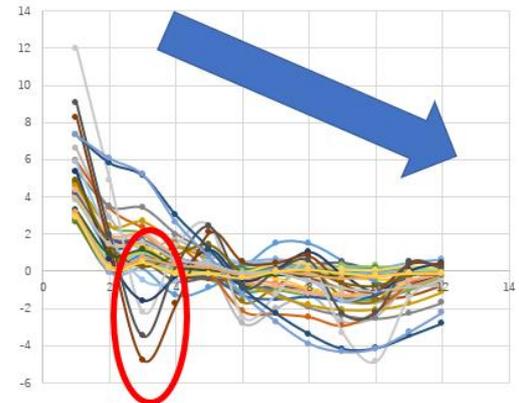


図6 リュウキュウキジバトのMFCC係数

Figure 6 MFCC coefficient of rufous turtledoves.

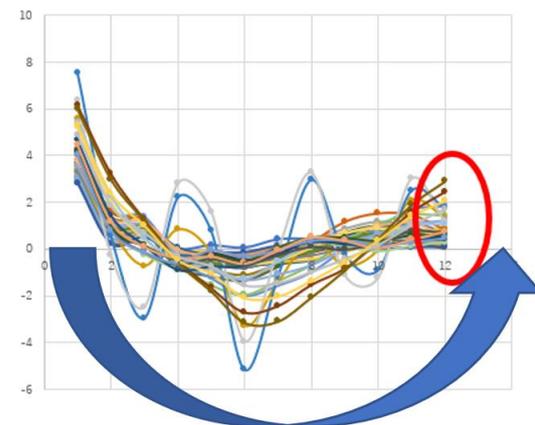


図7 ズアカアオバトのMFCC係数

Figure 7 MFCC coefficient for Zuaka green pigeon.

### 3.3 アルゴリズム

実験①、実験②-1については、野鳥のスペクトログラム画像を分類するモデルとして畳み込みニューラルネットワーク(図8)を実装するが、以下それぞれについて説明する。

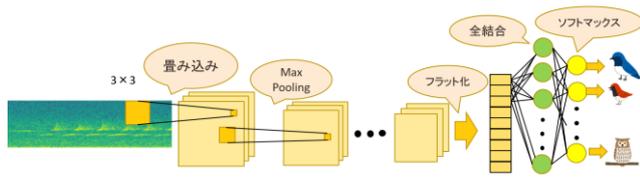


図 8 畳み込みニューラルネットワーク

Figure 8 convolutional neural network.

実験①と実験②-1 で構築するニューラルネットワークは入力画像が異なるだけで同じものを使用した。

では、具体的な中身について説明する。

まず画像を入力し、畳み込みを行う。畳み込みでは、Keras の Conv2D を使用し、3×3 のフィルタを 64 枚使用している。入力画像は、縦 64 横 64 ピクセルのカラー 3 色としている。そしてフィルタ後の画像に活性化関数「ReLU」を実施する。

次に、プーリングについて説明する。2×2 の大きさの最大プーリング層を作成する。ここでは 2×2 の領域で最大の数値を出力するようにする。この最大プーリング層では単純にカーネルの中の最大値を特徴として抽出している。それに畳み込み演算を行う。上記同様、3×3 のフィルタを 64 枚使用している。それに活性化関数を適用し、もう一度プーリングを行う。

上記、畳み込みとプーリングが終わった後に画像データを列ベクトルにフラット化する。全結合では出力を 256 にした。全結合層の出力の次元は実験①では 4、実験②-1 では 3 とし、ソフトマックス関数を用いて、各出力の予想確率を計算させている。

訓練用の画像として、野鳥 1 種につき 30 枚、検証用の画像として、野鳥 1 種につき 10 枚を使用している。本実験ではホールドアウト法を用いており、50 枚のデータを、訓練 30 枚、検証 10 枚、テスト 10 枚と分けている。

教師あり学習であるので、ラベル付けが必要となるが、今回どの音声かどの鳥に該当するかは共同研究者の榮村助教がラベル付けしてくれたデータを使用した。

バッチサイズは 10 で、エポック数は 20 である。

分類する際のテストデータは上記の通り 10 枚で、これをもとに、識別率を算出した。

実験②-1 に関しては、後述の通り識別率が完全ではなかったもので、訓練データ、検証データ、テストデータを入れ替え、計 5 回分の結果について算出している。

SVM については、メルケプストラム (MFCC) 係数 12 次元の値を特徴とし、線形手法を用いて分類を行った。

識別率の算出方法であるが、訓練データとして 40 枚用いて、テストデータとして 10 枚用いている。これも実験②-1 と同様、訓練データとテストデータをそれぞれ入れ替え、5 回分の結果について算出するようにしている。

## 4. 実験結果

### 4.1 ①の実験結果

①については CNN の画像識別で 6 種から 4 種へと分類している。6 種とは、アカヒゲ、カラスバト、オオトラツグミ、ルリカケス、リュウキュウキジバト、ズアカアオバトである。これを、アカヒゲ、オオトラツグミ、ルリカケス、ハト類という 4 種へ分類している。以下の表 1 は、実験①についての識別結果を示している。

表 1 ①の実験結果

Table 1 Experimental results of ①.

野鳥	正解	識別率
アカヒゲ	アカヒゲ	100%
カラスバト	ハト類	100%
オオトラツグミ	オオトラツグミ	100%
ルリカケス	ルリカケス	100%
リュウキュウキジバト	ハト類	100%
ズアカアオバト	ハト類	100%

表 1 のように、すべての鳥について識別率 100% という結果を出すことができた。

### 4.2 ②-1 の実験結果

実験②-1 では、スペクトログラム画像を用いた画像認識を使ったハト類 3 種についての分類である

3 種のハトとは、カラスバト、リュウキュウキジバト、ズアカアオバトのことである。

以下表 2 ではその実験②-1 についての識別結果を示している。

表 2 ②-1 の実験結果

Table 2 Experimental results of ②-1.

野鳥	識別率(全 5 回)				
	カラスバト	100%	100%	100%	90%
リュウキュウキジバト	90%	90%	80%	100%	100%
ズアカアオバト	100%	100%	100%	100%	100%

表 2 から、ズアカアオバトについては、5 回とも 100% という精度を得ることができた。カラスバトについては、4 回は 100% と出たものの、1 回のみ 90% となった。リュウキュウキジバトについては 1 回 100% と出て、あとは 80% が 2 回、90% が 2 回という結果になった。

それぞれ平均すると、カラスバトが 98%、リュウキュウキジバトが 88%、ズアカアオバトが 100% である。

誤識別しているものは、カラスバトとリュウキュウキジバトであるが、間違えとして出ている結果はこの相互間での間違えであった。

### 4.3 ②-2の実験結果

5回分の結果については前述のとおり、訓練データとテストデータを入れ替えて算出したものである。

表3から、全体の識別率は70%~90%であることがわかった。平均すると、80.6%である。

各種野鳥についても平均すると、カラスバトが72%、リュウキュウキジバトが74%、ズアカオバトが96%である。

表3 ②-2の実験結果  
 Table 3 Experimental results of ②-2.

回数	全体	カラスバト	リュウキュウキジバト	ズアカオバト
1回	90%	70%	100%	100%
2回	80%	90%	50%	100%
3回	80%	60%	80%	100%
4回	83%	90%	70%	90%
5回	70%	50%	70%	90%

追加で、線形SVMであるlinearと非線形カーネルの一種であるRBFカーネルを用いて、比較した。結果としては、表4より、linearのほう90%で、86.7%となったRBFカーネルの方が良い結果となった。

表4 線形手法と非線形手法の比較

Table 4 Comparison of linear and nonlinear techniques.

linear 識別率	RBFカーネル 識別率
90%	86.7%

## 5. 考察

実験①については、第5章実験結果より、すべて識別率が100%と出ていることが分かった。これは、ハト類という似ているスペクトログラム画像を1まとめたことでのこのようなわかりやすい結果となったものだと考える。

一方、ハトを詳細にして、当初のとおり6種の分類を実行すると、カラスバトが10%、リュウキュウキジバトが30%、そしてズアカオバトが70%と識別率がかなり低く出てしまった。

また、カラスバトの音声ファイルでアカヒゲが背景で鳴いていると、カラスバトの音声ファイルをアカヒゲと認識することがあった。逆もまたしかりである。これはスペクトログラム画像の帯域がかなり広いため、ハト類の周波数帯域以外の特徴に反応してしまい、別の野鳥に識別されてしまうためである。ハト類のスペクトログラム画像はおおよそ1kHz未満のところしか使用しないため、それより高い周波数帯域で声紋特徴が出てしまうと、それがハト類を分類するうえで障害となってしまうのである。このようにハト

類をまとめないことに対する弊害は大きかった。

実験②-1について、ズアカオバトについては識別率100%、カラスバト、リュウキュウキジバトについても識別率80%以上の精度を出すことができた。ズアカオバトについて、すべて100%と出ている理由であるが、これはスペクトログラム画像の中の声紋がカラスバトやリュウキュウキジバトよりも少し高い位置に存在するためであると考えられる。

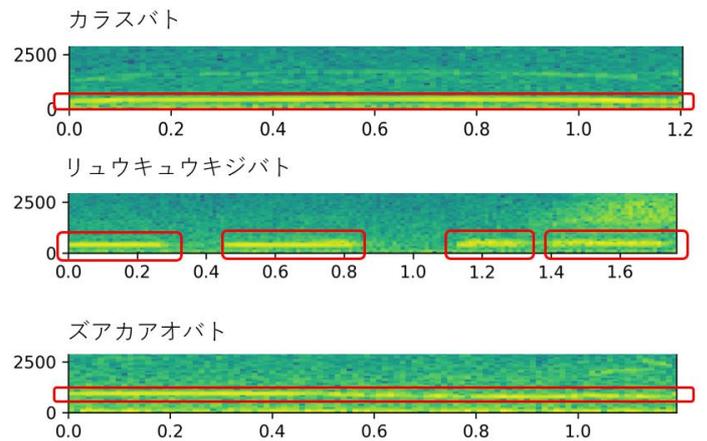


図9 ハト類のスペクトログラム画像比較

Figure 9 Comparison of spectrogram images of pigeons.

図9からも、ズアカオバトについては上2つよりも少し高い位置にあることが読み取れる。

そして、カラスバトとリュウキュウキジバトについて相互間で間違えがあるという理由についてだが、これはなぜだろうか。

たしかに、リュウキュウキジバトについては「ホーホーホー」という鳴き方をし、区切れがあるのが特徴である。ほとんどの音声ファイルではこのような鳴き方をしているのだが、ときよりその区切れが少ないファイルが存在する。以下図10のようなときである。

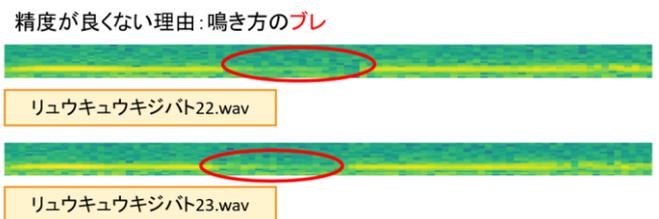


図10 リュウキュウキジバトについて鳴き方のブレ  
 Figure 10 The blurring of the ringing about the rufous turtledoves.

実験②-2について、実験結果より、ズアカオバトについては識別率90%以上、カラスバト、リュウキュウキジバトについては識別率50%~100%とまちまちだったがある程度の精度を出すことができた。全体としては70%~90%の識

別率だった。

実験②-1と同様に、ズアカオバトについては高い結果が出た。これは図7のように、カラスバト、リュウキュウキジバトに比べて、ズアカオバトがメルケプストラム(MFCC)係数にわかりやすく、U字のような特徴が出ていたからであると考えられる。図5,6より、カラスバトとリュウキュウキジバトのメルケプストラム(MFCC)係数が似ているため、やはりそんなに高い精度は得られなかった。

以下図11は、主成分分析により圧縮したメルケプストラム(MFCC)係数を使ったSVMによるハト3種分類である。画像の0,1,2はそれぞれ順にカラスバト、リュウキュウキジバト、ズアカオバトである。圧縮では、12次元の係数を2次元まで削減している。結果としてやはり圧縮すると、90%から66.7%まで識別率は落ちてしまった。しかし、この画像からも、0と1、つまりカラスバトとリュウキュウキジバトについての境界がうまく引けておらず、誤識別していることが読み取れる。

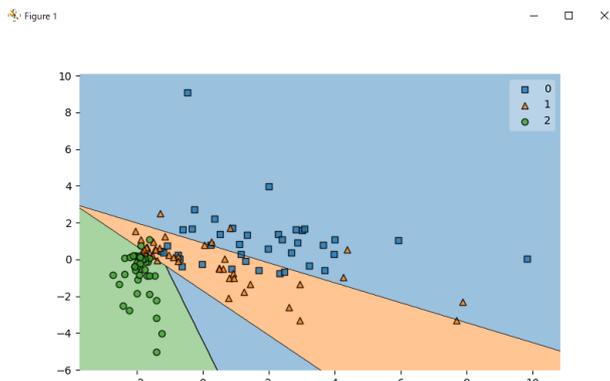


図 11 主成分分析によって圧縮した MFCC 係数  
Figure 11 MFCC coefficients compressed by principal component analysis.

他に、全体として精度が悪かった理由として、雑音の存在が挙げられる。

図12のように、リュウキュウキジバトのファイルにカラスの鳴き声が含まれていた。メルケプストラム(MFCC)係数は音色を特徴とする値であるため、そのような雑音には敏感に反応してしまうものだと考える。



図 12 雑音が入ったファイル 1  
Figure 12 File 1 with noise.

中には図13のように、カラスによる雑音が入っているが、あまり声紋として目立っていないものも存在する。



図 13 雑音が入ったファイル 1  
Figure 13 File 2 with noise.

距離が遠いからなのかよくわからないところではあるが、このようなところがスペクトログラム画像による画像認識と音色を特徴とするメルケプストラム(MFCC)係数で、同じファイルなのに識別結果が異なる理由の1つなのかもしれないと考えた。

## 6. まとめと今後の課題

本研究で識別対象としていた6種の野鳥については、高い精度で認識できていることがわかった。

そして、本研究では鳴き声等の特徴が似ている鳥についても識別を可能にするということを目指していた。実験では識別モデル2段階にすることにより、はじめに大まかな鳥の分類をし、次に似た鳥の分類を行うことでうまくいくことが分かった。段階を踏む際には、違う特徴を用いたり、画像の着目点を拡大するなどその鳥において認識できるような特化させたりすることが重要であることが分かった。同様の手順を踏むことで、識別対象が増えて、似ている鳥がいたとしても、細かい識別が可能になると予想する。

しかし、今回のハトは図9のようにスペクトログラム画像の使用帯域が近かったほか、特徴がわかりやすくでいたため、画像による識別が可能だった。よって、似ている鳥同士について、スペクトログラム画像で特徴がわかりやすくでない場合、メルケプストラム(MFCC)係数のような音色による特徴などを使い、適宜、分類器を作って識別していくべきであると考えられる。

現段階では6種についてしか分類していないが、目標の9種まで数を増やして再度実験に臨む必要がある。

また、本研究では使用したデータ1種の野鳥につき50データと少なかつたため、数百のデータに置き換えてから再度実験し、結果を出す必要がある。

それ以外に今後取り組むべきこととして以下をあげる。

第一に、判定する際、同一音声ファイル内に野鳥が重複していた場合に行う処理を考えることである。

第二に、自然音から野鳥の音を抽出する作業の確立である。本実験ではデータありきの実験だったため、環境音から識別を行うことができない。リアルタイムで観測できるようにするためにも、そのような処理をできるようにすることが求められる。

第三に、野鳥以外の鳥についての判定を行えるようにすることである。現在、どの野鳥に分類されるかという判定方法を用いているため、野鳥以外の動物の鳴き声が入った時の処理を考える必要がある。

## 参考文献

### 参考文献

- [1] 岩崎祐介, 三田長久, 牧野洋平, 高橋幸司, カムケオシーパーチャン, “音声情報を用いた夜行性鳥類の種識別”, FIT2007(第6回情報科学技術フォーラム)p329-330
- [2] 東谷幸治, 三田長久, 牧野洋平, “音声情報によるニューラルネットワークを用いた夜行性野鳥の識別”, FIT2006(第5回情報科学技術フォーラム)p355-356
- [3] NPO 法人奄美野鳥の会, 鳥飼久裕・水田拓, 島隆穂, 鈴木真理子, “世界自然遺産登録地奄美大島における録音装置を用いた鳥類モニタリング手法の開発と次世代の調査従事者の育成” 自然保護助成基金助成成果報告書 vol. 31(2022)p193-201
- [4] 斎藤有杜, 三上剛, “環境音からの野鳥の鳴き声を検出する畳み込みニューラルネットワークの構築”, 情報処理学会第84回全国大会 p2-19~2-20
- [5] 太田望, 今井俊輔, 延原肇, 徳江義宏, 今村史子, 太田敬一, 夏川遼, “深層学習による野外録音音声からの動物の鳴き声検出とその環境影響評価への応用-オオタカの鳴き声を事例として”, 情報処理学会第80回全国大会 p2-379~2-380
- [6] 東谷幸治, 三田長久, 他, 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, 電子情報通信学会総合大会講演論文集 2007 年情報・システム(1), 146, 2007-03-07
- [7] Seppo Fagerlund, Automatic Recognition of Bird Species by Their Sounds, Helsinki University of Technology, 2004-11
- [8] 窪田伊織, 丸山和暉, 中嶋響, 山崎皓平, 村井亮介, “鳥類音声録音装置の試作と自動種数数え上げシステム構築の試み”, 高知工科大学紀要 18-1 p41-49, 2021.12.28
- [9] 山本修生, 伊藤克亘, “ニューラルネットワークによる実環境下での環境音認識”, 情報処理学会第79回全国大会, p3-445~3-446
- [10] 高橋栄作, “津軽弁の音声認識” 「地域政策研究会」(高崎経済大学地域政策学会)第23巻 第4号 p63-74 2021.3
- [11] 曾我真子, 入部百合絵, 兒嶋朋貴, 増田達明, 愛知県立大学情報科学部, 愛知県農業総合試験場, “個体識別を目的とした牛音声の音響情報分析”, 情報処理学会第81回全国大会 p2-561~2-562
- [12] スマホやPCで鳴き声をAIで解析して鳥の種類を判別できる「BirdNET」が登場  
<https://gigazine.net/news/20210802-birdnet/>, (参照 2023-02-18)
- [13] LibROSA で MFCC (メル周波数ケプストラム係数) を算出して楽器の音色を分析  
<https://www.wizard-notes.com/entry/music-analysis/insts-timbre-with-mfcc/>, (参照 2023-02-18)
- [14] 音声特徴量 (ケプストラム, メル周波数ケプストラム

係数) の理解に役に立つ, 声の生成の仕組み

[https://qiita.com/tmtakashi\\_dist/items/eecb705ea48260db0b62](https://qiita.com/tmtakashi_dist/items/eecb705ea48260db0b62), (参照 2023-02-18)

[15] Q 様々な音響特徴量 それぞれの使い方や意味を教えてください 東北大学大学院工学研究科博士後期課程2年千葉祐弥 [http://abcpedia.acoustics.jp/acoustic\\_feature\\_2.pdf](http://abcpedia.acoustics.jp/acoustic_feature_2.pdf), (参照 2023-02-18)