

平成 29 年度 特別研究報告書

**鳴き声を用いた鳥の自動分類における
最適パラメータの検討**

龍谷大学 理工学部 情報メディア学科

T140501 前川 志帆

指導教員 三好 力 教授

目次

内容梗概.....	4
第1章 緒言.....	5
第2章 研究背景.....	6
2.1 野生動物生息調査の現状.....	6
2.1.1 一般的な調査方法.....	6
2.1.2 野鳥における調査方法.....	6
2.2 同種研究の現状.....	7
2.3 本研究の目的.....	7
第3章 新しいパラメータの試案.....	8
3.1 人と鳥の出せる音.....	8
3.2 既存の特徴量算出方法.....	8
3.2.1 メル周波数分析.....	9
3.2.2 ケプストラム分析.....	10
3.3 本研究における MFCC の問題点.....	10
3.4 新しいパラメータの試案.....	11
3.5 Bird' s MFCC の算出.....	12
3.5.1 算出方法.....	12
3.5.2 算出結果.....	13
第4章 評価検討実験.....	14
4.1 評価方法.....	14
4.2 実験環境.....	14
4.3 実験データ.....	14
4.4 実験結果と考察.....	16
第5章 結論.....	20
謝辞.....	21
参考文献.....	22

付録..... 23

内容梗概

野鳥の調査は、主に捕縛か観測と鳴き声の聞き分けによって行われる。野鳥の多いイギリスでは、計算機を用いてこの鳴き声による調査を簡単に行えるようにする研究が注目されている。楽曲等の分類に比べてその時々で変化する複雑な野鳥の声を分類する為、研究の事例がまだ少ない。

本研究では、鳴き声による野鳥の自動分類を考えて、鳥の鳴き声の特徴をよく表わすパラメータをつくることを試みている。人の声道・聴覚特性を表わすパラメータ「MFCC」を基本として、メルフィルタバンクをかけ合わせる処理に工夫をし、鳥の鳴き声に合わせて高い音の違いを細かく表す「Bird's MFCC」としている。

また、その Bird's MFCC がより鳥の鳴き声に合わせてできているのかを調べる為、 k - means 法を用いて 3 種類の鳥を分類する。その分類精度等を比較検討して、Bird's MFCC のパラメータとしての妥当性を検討している。

MFCC の分類の正答率は 73.3%、Bird's MFCC の分類の正答率は 96%と、正答率は MFCC を大きく上回る結果となった。また、不正解データが、MFCC は高さの大きく異なるクラスに含まれるのに対して、Bird's MFCC では高さの近いクラスに分類されている。これらの事から、MFCC と Bird's MFCC は異なる特徴を表わし、Bird's MFCC の方がより鳥の鳴き声の特徴を表わすことが分かる。

第1章 緒言

一般的に動物の生息調査には、糞やその足跡等の痕跡を利用する。これは、動物たちにストレスを与えず、捕縛の難しい種でも簡単に行えるためである。鳥においては、その鳴き声から種や数を推定している。これを計算機によって行うことができれば、より容易に鳥の生息調査を行う事ができる。その為、鳴き声から鳥の種類を判別する研究[1]が行われている。しかし、楽曲の推定とは異なり、同じ種の個体でも様々な鳴き方をする鳥の鳴き声での判別は容易ではなく、あまり研究の事例がない。また、先行研究[1]では、機械学習の過程に注目しており、そのパラメータの検討においては不十分である。そこで、より鳥の声に対応したパラメータとして、鳥の鳴き声の高さに合わせて、その高さにおいて細かく違いを表わすことができるパラメータを考える。

本研究では、高鳥の鳴き声に合わせて、高音域の音の特徴をより細かく表すパラメータ「Bird's MFCC」を提案する。これは、人間の聴覚特性や声道特性に合わせた特徴量 MFCC を参考にしている。MFCC では、人間の聴覚特性に合わせる為、低音域の音の特徴をより細かく表すために、メルフィルタバンクという対数フィルタを掛け合わせて低音域での違いを強調している。このメルフィルタバンクの強調域を高音域に合わせることで、鳥の鳴き声に合わせて、その差を強調する事ができる特徴量「Bird's MFCC」とする。また、このパラメータが正しく鳥の鳴き声の特徴を強調しているのかを調査する為、3種類の鳥の鳴き声から MFCC と Bird's MFCC を算出し、*k*-means 法を用いてそれぞれを 3 クラスに分類した結果を比較し、この特徴量の妥当性を検討する。

本論文では、まず、第 2 章で本研究の前提となる野生動物生息調査の方法や同種研究の現状について述べ、本研究の目的について述べる。第 3 章では、まず、人間と鳥での発声に関する違いを述べ、本研究で提案する Bird's MFCC の元となる MFCC の処理手順や本研究における問題点を述べ、その改善方法について述べる。第 4 章では、MFCC と Bird's MFCC での分類結果について述べ、Bird's MFCC のパラメータとしての妥当性について述べる。最後に第 5 章では、本論文をまとめ、今後の課題について述べる。

第2章 研究背景

2.1 野生動物生息調査の現状

2.1.1 一般的な調査方法

どの場所においても、野性動物は存在する。その動物が地域環境に与える影響や地域動物の生態系との関連などを調べるために、各地で生息状況の調査が行われている。その中で、基礎的な情報として、生息する種の特異とその種の生息数の推定がある。これらは、実際に現地に向かい、観測調査を行っている。例えば、ノウサギの調査[2]に関しては、その足跡から数を推定する INTGEP 法や空中センサス法、糞から数や種を推定する糞粒法、捕縛して数や種を特定する追い出し法などが用いられている。この捕縛により数や種を特定する方法は、生息調査において一番確実な方法として挙げられるが、動物にストレスを与える為、調査後に死に至る動物も少なくない。その為、貴重種の調査には不向きである。また、鳥や大型動物など捕縛自体が難しい種にもこの調査方法は適さない。その為、足跡や糞による推定調査が現在の野生動物生息調査における主な調査方法となっている。

2.1.2 野鳥における調査方法

2.1.1 で述べたように、野生動物の生息調査において、足跡や糞による調査が一般的であるが、野鳥は足跡を残すことが少なく、糞の場所も特定されにくい為、これらの調査方法では調査できない。その為、野鳥においては、目視と鳴き声の聞き分けによって種や数を調査するラインセンサス法やスポットセンサス法[3]が主である。目視に関しては写真での記録も可能な為、誰でも行う事ができるが、鳴き声のみの聞き分けになると専門的な知識が必要となり、専門家が行うこととなる。野鳥調査において先行的であるイギリスでは、この鳴き声による鳥の分類を、計算機を用いることで自動化する研究を行っている。

2.2 同種研究の現状

特に野鳥の多いイギリスでは, 野鳥調査の自動化に注目が向けられている. 同国の Florence Wilkinson, Dan Stowell らによる Warblr チーム[4]は, ロンドン大学のクイーンメアリー校等の支援を受け, イギリスの野鳥の鳴き声を学習し分類するアプリ Warblr[1]を開発している. しかし, 世界的に見ると, このような野生動物の生息調査を助ける研究はまだ少なく, Warblr チームが行ったクラウドファウンディングも目標に届かず, まだ開発途中とされている.

2.3 本研究の目的

そこで, 本研究では, 野鳥の鳴き声による生息調査に注目し, 計算機による自動化を考え, 鳥の鳴き声の特徴をより細かく表す新しいパラメータを提案することを目的とする.

第3章 新しいパラメータの試案

3.1 人と鳥の出せる音

鳥の鳴き声の特徴を考える上で、人との違いを考える。一般的に、人間が出せる音は、80Hz から 3kHz, 可聴域は 20Hz から 20kHz とされている。しかし、表1に表わされるように、鳥は比較的低い音を出すミゾゴイやツツドリで 300Hz, 高い音を出すクロツグミやヤブサメで 8kHz 辺りの音を出すとされており、鳥は人間に比べて高い音を出すと云える[5]。

表 1：高さの異なる主な鳥の鳴き声

鳥の種類	低い鳴き声(Hz)	高い鳴き声(Hz)
ミゾゴイ		329.63
フクロウ		415.30
カッコウ	698.46	880.00
ホトトギス		1760.00
コチドリ	2489.02	3135.96
メジロ		4186.01
エゾムシクイ	5274.04	8372.02
カワセミ		5919.91
ヤブサメ		8372.02

3.2 既存の特徴量算出方法

音声の分析には、主に MFCC(Mel-Frequency Cepstrum Coefficients)[5]が用いられる。MFCC は、メル周波数分析とケプストラム分析を合わせたもので、人間の聴覚特性と声道特性に合わせて特徴をより分かりやすく表した特徴量、また、その特徴量計算の手法である。その処理の手順を、図 1 に示す。

図 1 より、まず音声を入力し、その音声にハン窓をかける。ハン窓をかけた音声に短時間フーリエ変換(STFT : Short-Term Furier Transform)を行う。変換後の振幅スペクトルにメルフィルタバンクをかけ合わせ、スムージングを行う。そうして得られた離散信号に離散コサイン変換(DCT : Discrete Cosine Transform)を行うと MFCC を得ることができる。

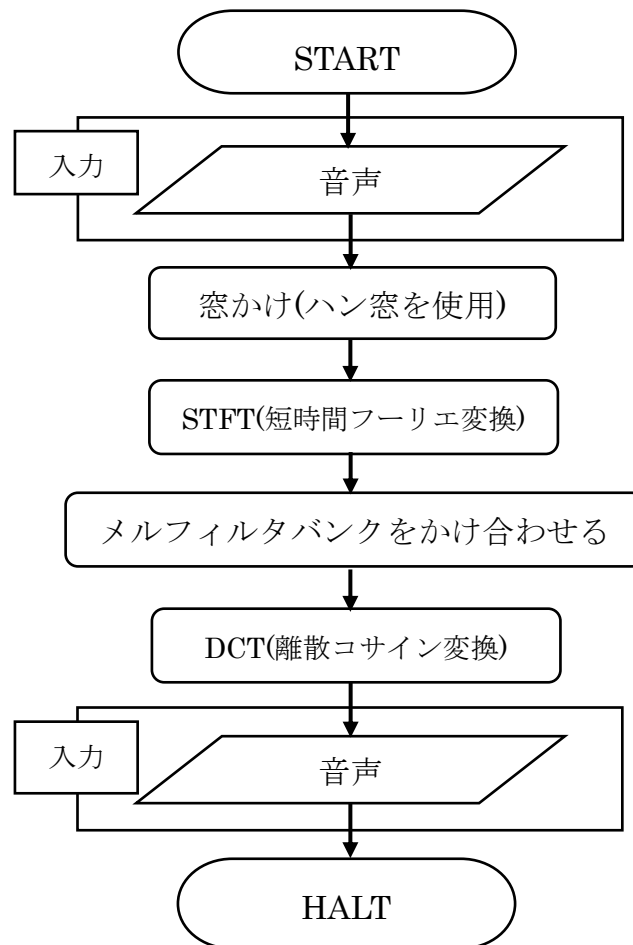


図 1 : MFCC の処理フロー

3.2.1 メル周波数分析

図 1 の MFCC の処理のうち、メルフィルタバンクを掛け合わせる処理をメル周波数処理と言い、人間の音高の知覚特性(可聴域の下限(20Hz)に近い音は高めに、可聴域の上限(20kHz)に近い音は低めに聞こえる特性)や聴覚特性(低い音ほどより細かな違いを聞き分けられる特性)に合わせる処理である。

メルフィルタバンクというのは、人間の聴覚特性に合わせたメルスケール(図 2)に沿って等間隔に配置された三角関数のフィルタであり、これを掛け合わせることで、低い音ほどより細かく特徴が表れ、周波数方向に平均化される。

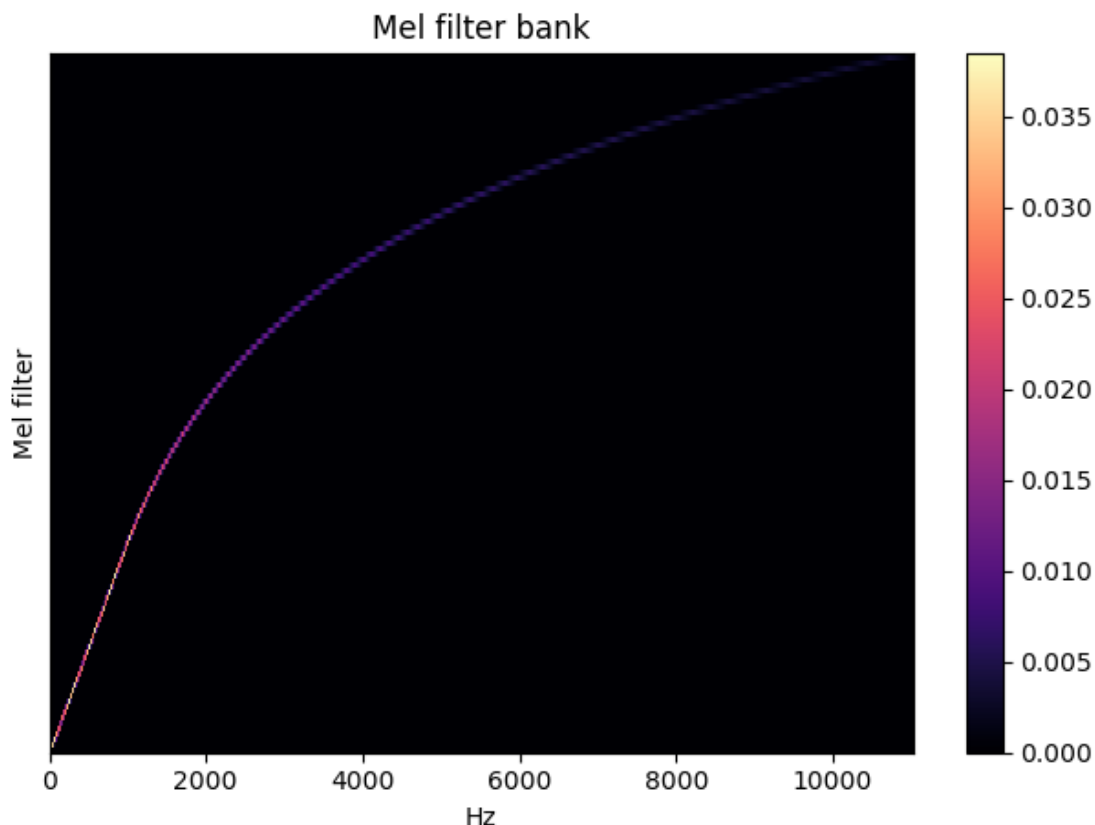


図 2：メルスケール

3.2.2 ケプストラム分析

一般的に、ケプストラム分析は、音声信号にフーリエ変換を行い、その周波数スペクトルの対数を取ったものに逆フーリエ変換を行うことを示す。MFCC では、対数を取る代わりにメルフィルタバンクをかけ合わせ、逆フーリエ変換の代わりに離散コサイン変換を行っている。ケプストラム分析を行うことで、人間の聴覚特性に合わせたスペクトル微細構造(周波数方向の細かな変動特徴)と人間の声道特性に合わせたスペクトル包絡(周波数方向のなだらかな変動)を分離して表わすことができる。

3.3 本研究における MFCC の問題点

図 2 から分かるように、人間に合わせたメルスケールでは、特に 0Hz から 2000Hz の音がより細かく表されるようになっている。表 1 で確認したように、鳥にはこれより高い音しか出すことができない種も少なくない。その為、メルスケールの強調周波数を工夫する必要がある。

3.4 新しいパラメータの試案

鳥は人間に比べて高い音を出す種が多いため、その高音域の違いをより細かく表せるように考える。MFCC ではメルスケールを用いて、低音域を強調している。これに着目し、メルスケールの対象音域を上げることで高音域の特徴をより細かく表す特徴量 Bird's MFCC とする。具体的には、図 3 に示すように対数関数の最小値を引き上げる。こうすることで、鳴き声とは関係のない低音域の雑音が含まれた場合、それらを加味せずに特徴量が求められるようになる。

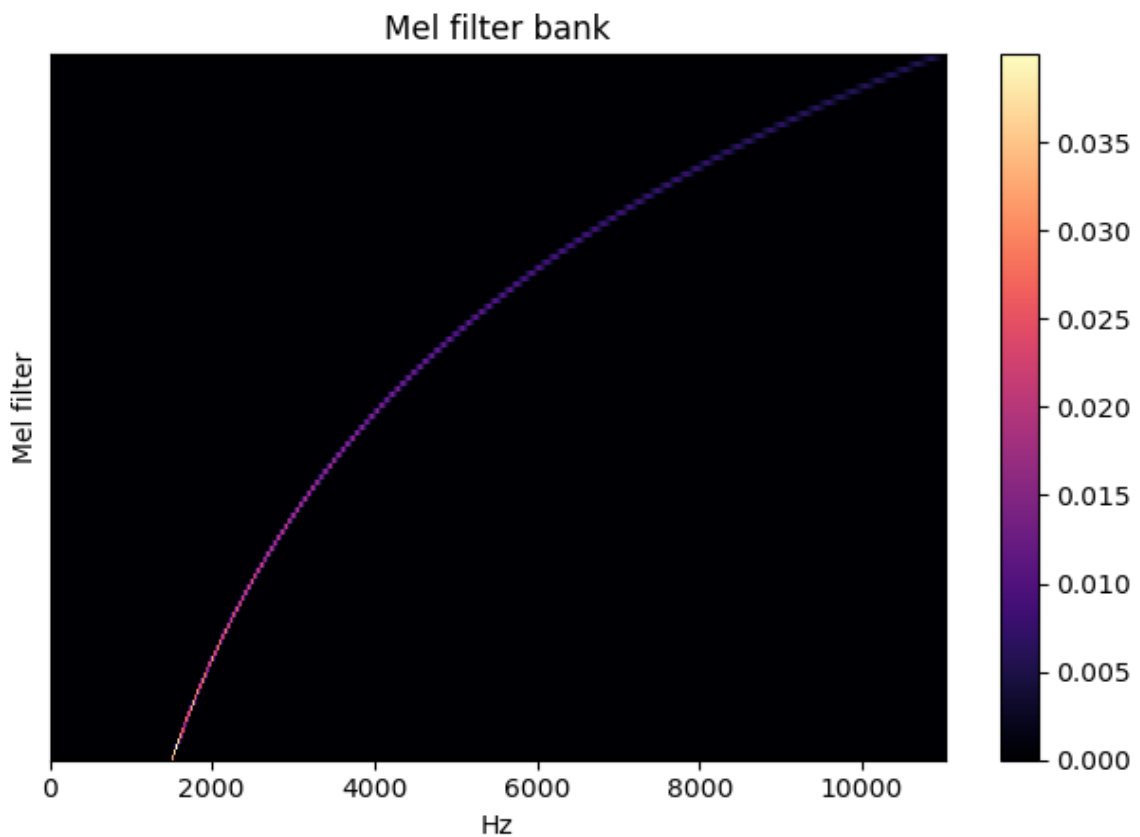


図 3 : 強調域を 1500Hz から 3500Hz 辺りとするメルスケール

3.5 Bird's MFCC の算出

3.5.1 算出方法

実際に変更したメルスケールを用いて、Bird's MFCC を算出する。使用した音源は、Long-tailed cuckoo の鳴き声 1 回分が収録されている 5 秒の WAVE 形式のファイルである。Python の音楽分析用パッケージである librosa の関数を用いて MFCC を算出する。MFCC 関数を使用すると、メルスケールは図 2 のように 0Hz から 2000Hz 辺りを強調し、13 次元のベクトルとしてあらわされる(図 4)。その為、図 1 のフローに従ってそれぞれの処理を行い、メルスケールの強調周波数域を図 3 のように 1500Hz から 3500Hz 辺りを強調するように変更すると、Bird's MFCC が得られる(図 5)。

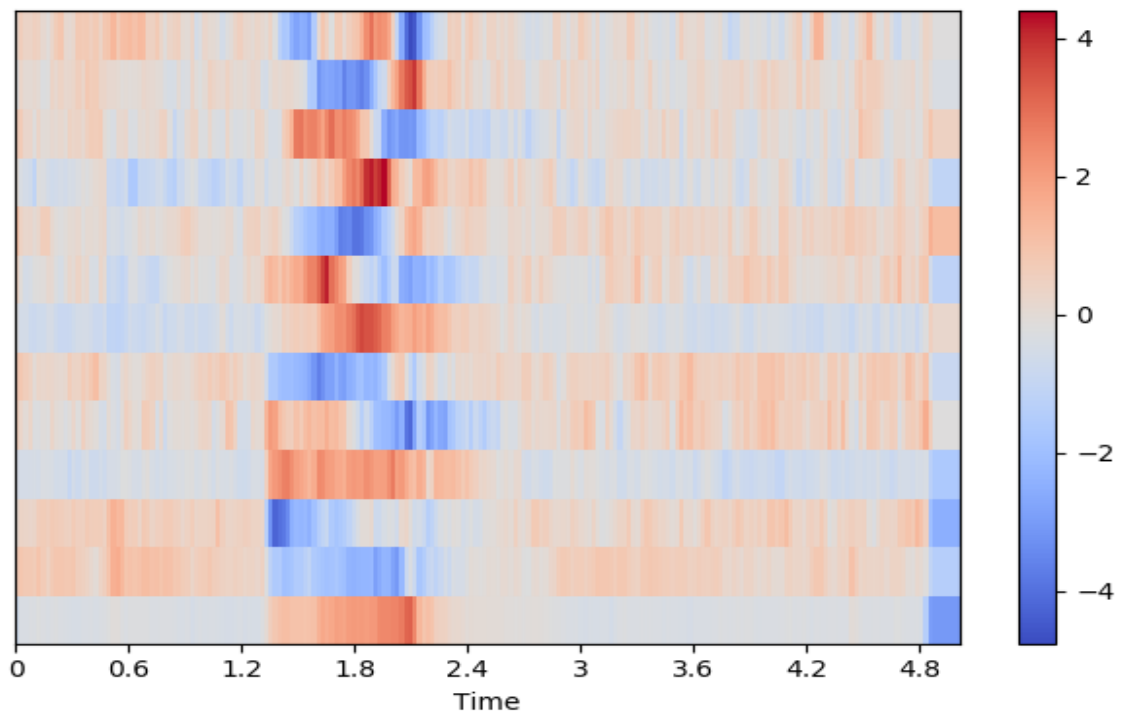


図 4 : Long-tailed cuckoo の MFCC

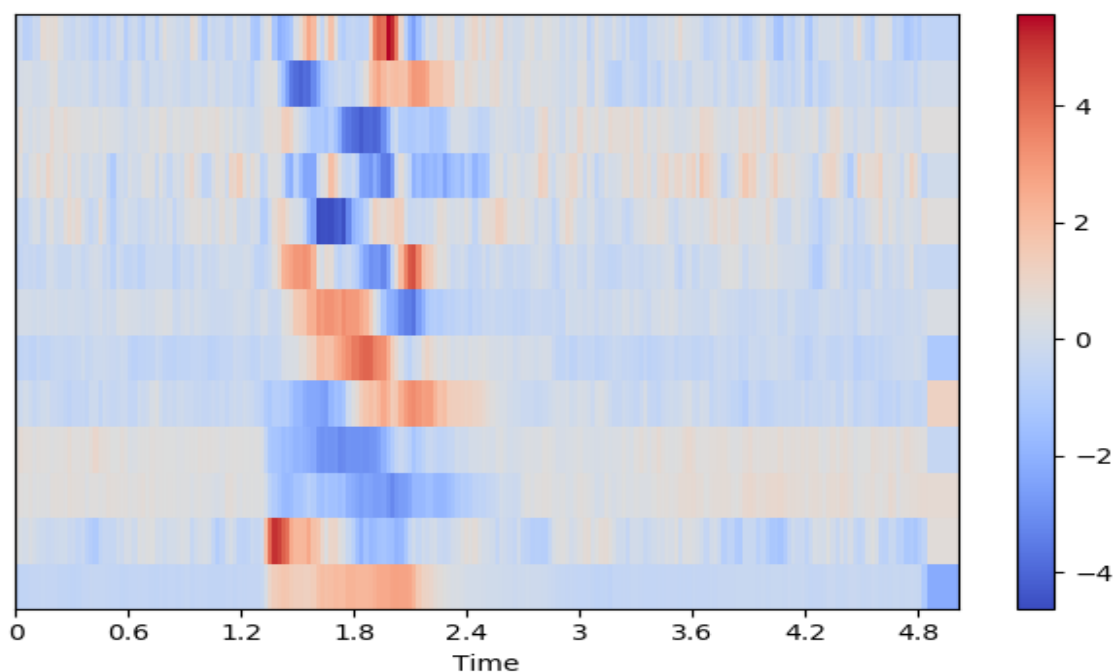


図 5 : Long-tailed cuckoo の Bird's MFCC

3.5.2 算出結果

図 4 と図 5 の全体の色から分かるように、雑音部分が低減されているのはわかる。また、図 4 の中で最も強い特徴を表わす赤い部分は、特徴を荒くする高次元のところにあるのに対して、図 5 の赤い部分は細かく表す低次元の部分にあることが分かる。

第 4 章 評価検討実験

4.1 評価方法

メルスケールの強調周波数域の変更によって、MFCC よりも Bird's MFCC の方が、より鳥の特徴を表わすことができているのかを考える。具体的な方法としては、非階層型クラスタリング手法の k -means 法を用いて、3 種類の鳥 (Long-tailed cuckoo : キジカッコウ, Kea : ミヤマオウム, Morepork : ニュージーランドアオバズク) の鳴き声の MFCC・Bird's MFCC を算出し、それぞれを 3 クラスに分類する。その分類精度等に関して比較検討を行う。

また、Bird's MFCC の強調周波数域についても最適な値を検討する為、メルスケールの対象の最低値が 0Hz である MFCC, 500Hz, 1000Hz, 1500Hz, 2000Hz である Bird's MFCC の合計 5 種類の特徴量を用いて、分類精度等に関して比較検討を行う。

4.2 実験環境

k -means 法を用いるために使用するのは、機械学習ソフトウェアの weka3.8.1 である。使用音源は Long-tailed cuckoo・Kea・Morepork の鳴き声 1 回分(2 秒から 5 秒) が収録されている WAVE 形式のファイルである。それらに対して、MFCC と Bird's MFCC を算出し、 k -means 法で 3 クラスにクラスタリングを行う。

4.3 実験データ

比較実験を行う前に、対象とする 3 種類の鳥の鳴き声に関して調査する。

まず、Long-tailed cuckoo : キジカッコウは「キューイ」というような音で、後半になるにつれて音が上がる鳴き方をする。Long-tailed cuckoo の鳴き声の STFT を算出した結果を図 6 に表す。図 6 から分かるように、主な鳴き声の高さは 2000Hz から 10000Hz である。

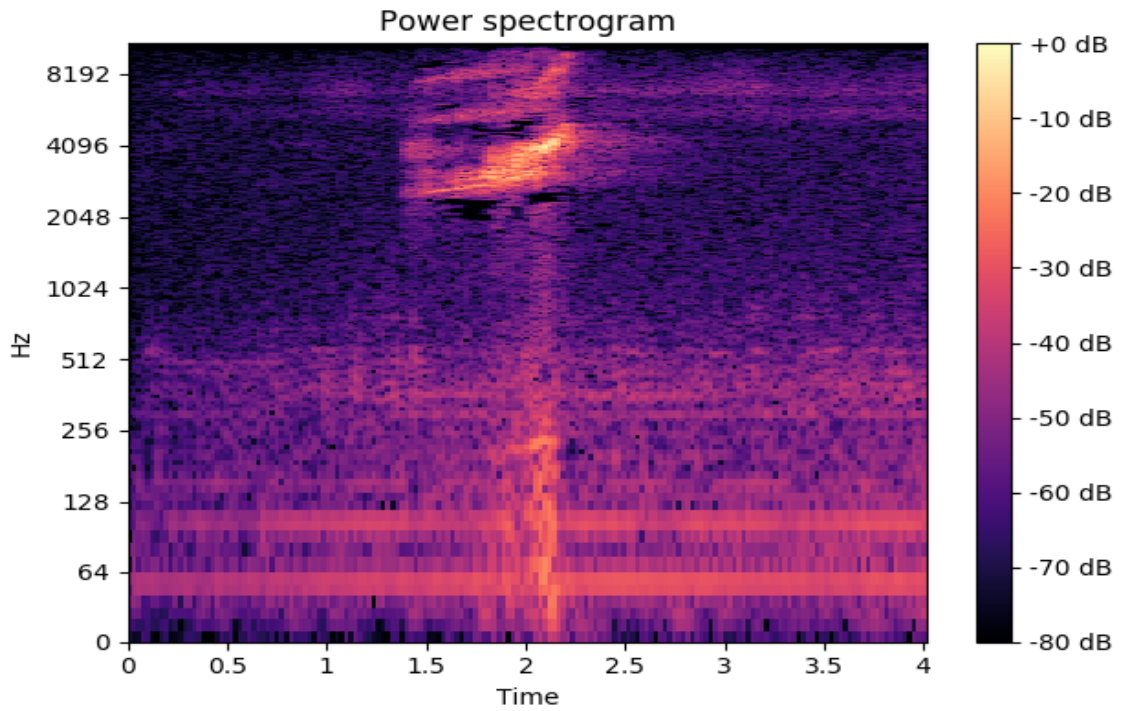


図 6 : Long-tailed cuckoo の STFT

次に、Kea : ミヤマオウムは「キアー」とビブラートにきいた、後半になるにつれて音が下がる鳴き方をする。Kea の鳴き声の STFT を算出した結果を図 7 に表す。図 7 から分かるように、主な鳴き声の高さは 1000Hz から 7000Hz である。

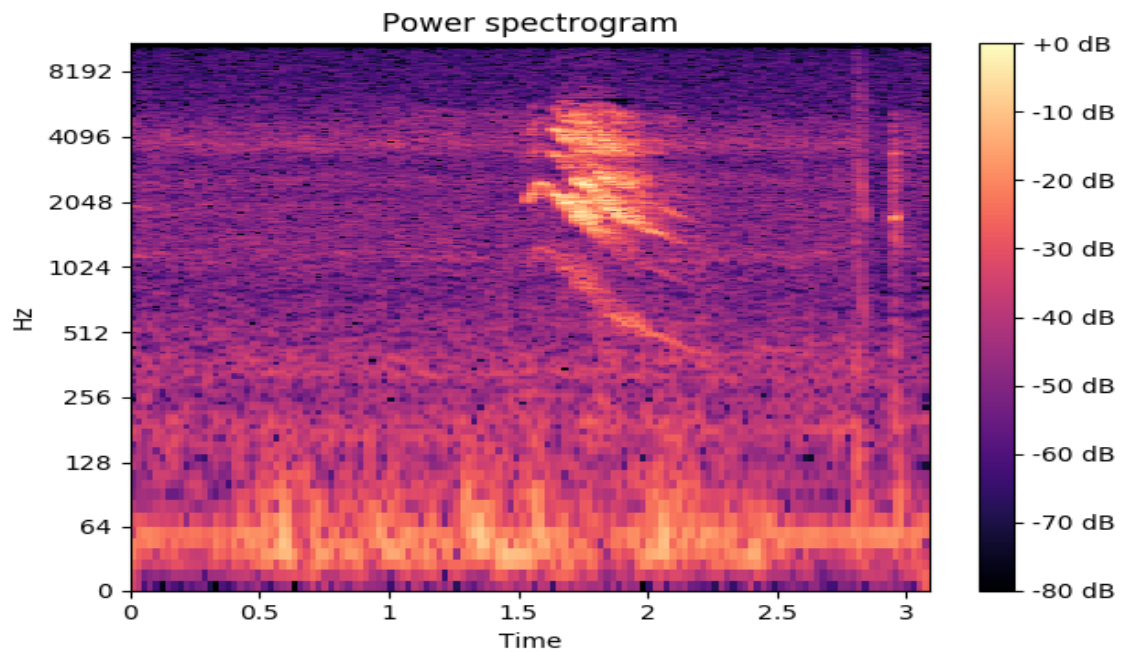


図 7 : Kea の STFT

最後の、Morepork : ニュージーランドアオバズクは「ホーホー」と短く連続で、フクロウ科にしては少し音の高い鳴き方をする。Morepork の鳴き声の STFT を算出した結果を図 8 に表す。図 8 から分かるように、主な鳴き声の高さは 700Hz から 1000Hz である。

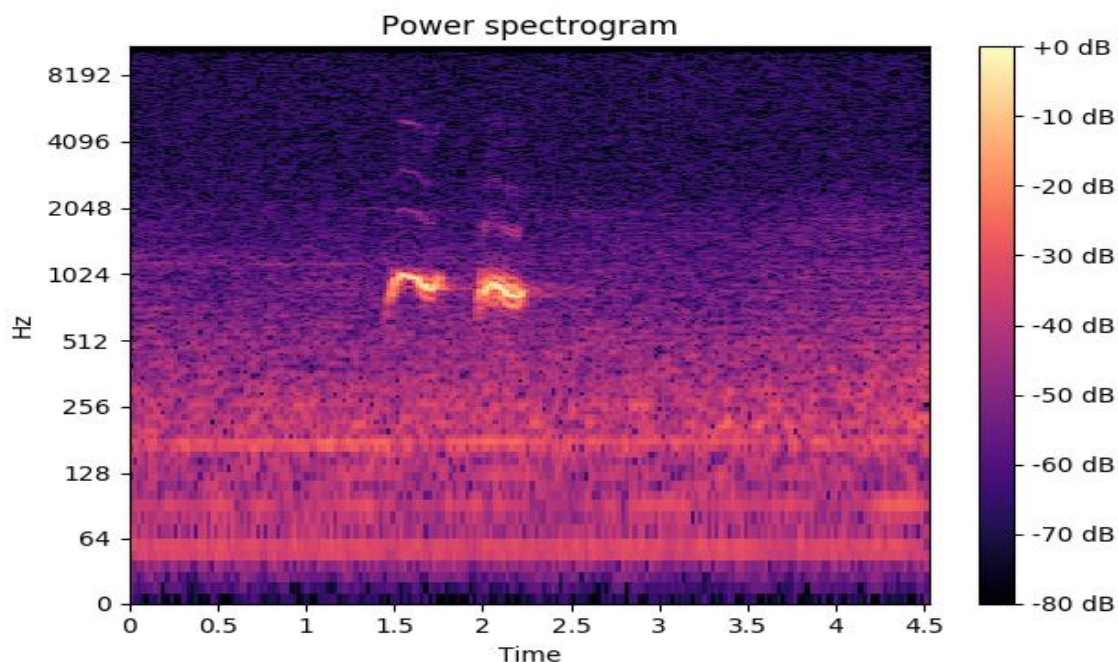


図 8 : Morepork の STFT

4.4 実験結果と考察

これら 3 種類の鳥の鳴き声に対して、MFCC と 4 つの Bird's MFCC を算出し、3 クラスに分類した結果を、次の表 2 から 6 にまとめる。列方向が正解とする鳥の名前、行方向が判定された鳥の名前、色のついた部分の値が正しく分類されたデータの個数である。また、5 つの特徴量に対して、それらの各正答率と総合正答率を表 7 にまとめる。

表 7 から、強調周波数域の最低値が 1000Hz の(1000Hz から 3000Hz 辺りの音がより強調される)Bird's MFCC の正答率が一番高いことが分かる。Kea・Morepork の分類においては、100%の正答率でクラスタリングが行われている。

また、総じて正答率の低い Long-tailed cuckoo の鳴き声は、不正解データが鳴き声の高さの近い Kea ではなく、鳴き声の高さの大きく異なる Morepork のクラスに偏っているものが MFCC を含めてほとんどである。しかし、最低値 1500Hz の Bird's MFCC では、Morepork との間違いはなく、MFCC とはクラスタリングの基準が異なることが分かる。

これらのことから、強調周波数域 1000Hz から 3000Hz 辺りの Bird's MFCC は、MFCC では表せなかった鳥の鳴き声の特徴を表わすことができる、ということが分かる。

表 2 : MFCC のクラスタリング結果

		正解		
		Long-tailed cuckoo	Kea	Morepork
判定	Long-tailed cuckoo	5	0	0
	Kea	0	7	0
	Morepork	5	3	10
正答率		50%	70%	100%

表 3 : 最低値 500Hz の Bird's MFCC のクラスタリング結果

		正解		
		Long-tailed cuckoo	Kea	Morepork
判定	Long-tailed cuckoo	5	0	0
	Kea	0	6	0
	Morepork	4	3	9
正答率		56%	67%	100%

表 4 : 最低値 1000Hz の Bird's MFCC のクラスタリング結果

		正解		
		Long-tailed cuckoo	Kea	Morepork
判定	Long-tailed cuckoo	7	0	0
	Kea	0	8	0
	Morepork	1	0	8
正答率		88%	100%	100%

表 5 : 最低値 1500Hz の Bird's MFCC のクラスタリング結果

		正解		
		Long-tailed cuckoo	Kea	Morepork
判定	Long-tailed cuckoo	5	2	0
	Kea	5	3	0
	Morepork	0	5	10
正答率		50%	30%	100%

表 6 : 最低値 2000Hz の Bird's MFCC のクラスタリング結果

		正解		
		Long-tailed cuckoo	Kea	Morepork
判定	Long-tailed cuckoo	4	0	0
	Kea	0	7	0
	Morepork	4	1	8
正答率		50%	88%	100%

表 7 : 全クラスタリング結果のまとめ

		正答率			
		Long-tailed cuckoo	Kea	Morepork	総合
使用した特徴量	MFCC	50%	70%	100%	73%
	最低値 500Hz の Bird's MFCC	56%	67%	100%	74%
	最低値 1000Hz の Bird's MFCC	88%	100%	100%	96%
	最低値 1500Hz の Bird's MFCC	50%	30%	100%	60%
	最低値 2000Hz の Bird's MFCC	50%	88%	100%	79%

第 5 章 結論

本研究では、野鳥の調査に役立てることを目的に、野鳥を鳴き声で分類するのに最適な特徴抽出法を検討した。

まず、人の声道・聴覚特性に合わせた MFCC に注目し、その中で人の聴覚特性に合わせる為に、低い音ほど細かい特徴を強調するメルフィルタバンクをかけ合わせる処理に注目した。このメルスケールの強調域を鳥に合わせて高くすることで、より鳥の鳴き声に合わせた特徴量を表わすことができると考え、**Bird's MFCC** とした。

更に、この **Bird's MFCC** が MFCC よりも鳥の鳴き声の特徴量をより表すことができているのかを確かめるために、*k* - means 法を用いて 3 種類の鳥の鳴き声からそれぞれを算出し、3 クラスに分類した。結果としては、強調周波数域を 1000Hz から 3000Hz とする **Bird's MFCC** の正答率が MFCC の正答率より高く、96% という結果となった。また、鳴き声の高さが大きく異なる種類と間違えていた MFCC に対して、**Bird's MFCC** では高さの似ている種類との間違いのみになっていた。このことから、MFCC では表すことができない特徴を **Bird's MFCC** では表わすことができる、ということが調べられた。

今後の課題としては、まずは今回の **Bird's MFCC** では低音域を切り捨てて高音域を強調していた為、低音域を含んだ上で高音域のみを強調する新しい **Bird's MFCC** が考えられる。様々な鳥に対応していく必要がある為、同時に扱う鳥の種類に関してもデータを変えていく必要があり、扱う鳥のデータをより高さが似ているものに変え、その高さに合わせて強調区域を絞ることで、より MFCC との差別化が図る事ができると考えられる。また、**Bird's MFCC** は主に高さを表わす特徴量である為、それ以外の要素にも着目できるよう、更に新しいパラメータを試案していく必要もあると考えられる。

謝辞

本研究を進めるにあたり暖かい御指導，御鞭撻を頂いた龍谷大学工学部情報メディア学科三好力教授に深く感謝いたします。そして，講義，実験等において御指導を頂いた 龍谷大学工学部情報メディア学科岡田至弘教授，片岡章俊教授，新川芳行教授，外村佳伸教授，野村竜也教授，長谷智弘教授，藤田和弘教授，曾我麻佐子准教授，吉見毅彦准教授，奥健太講師，渡辺靖彦講師，芝公仁助教，岩嶋浩樹実験講師，寄能雅文実験助手に厚く御礼申し上げます。

また，研究において多方面に渡り御支援下さった龍谷大学工学部情報メディア学科，及び龍谷大学工学研究科情報メディア学科専攻三好研究室の諸氏に厚く御礼申し上げます。

参考文献

- [1] warblr 社, "Warblr : the birdsong recognition app", <https://warblr.net/>, (参照 2017-8-16).
- [2] 環境省, "狩猟鳥獣のモニタリングのあり方検討会",
<https://www.env.go.jp/nature/choju/effort/effort6/effort6.html>, (参照 2017-7-27).
- [3] 環境省, "モニタリングサイト 1000",
<https://www.biodic.go.jp/moni1000/manual/index.html>, (参照 2017-7-27).
- [4] waikato 大学, "Weka 3 : Data Mining Software in Java",
<http://www.cs.waikato.ac.nz/ml/weka/>, (参照 2017-11-13).
- [5] "音階(周波数)と鳥の鳴き声",
<http://birds.oosato.org/cgi-bin/documentscgi?text=docWavScale>, (参照 2017-7-27)
- [6] Beth Logan, "Mel Frequency Cepstral Coefficients for Music Modeling",
http://ismir2000.ismir.net/papers/logan_paper.pdf, (参照 2017-8-20)

付録

メルスケールの描画プログラム

```
import matplotlib.pyplot as plt
import librosa.display

mel_fb = librosa.filters.mel(22050, 2048, fmin = 1500)

plt.figure()
librosa.display.specshow(mel_fb, x_axis='linear')
plt.ylabel('Mel filter')
plt.title('Mel filter bank')
plt.colorbar()
plt.tight_layout()
plt.show()
```

STFT の算出・描画プログラム

```
import wave
import numpy as np
import matplotlib.pyplot as plt
import librosa.display

music, fs = librosa.audio.load("Morepork1.wav")

D = librosa.stft(music)

librosa.display.specshow(librosa.logamplitude(np.abs(D)**2, ref_power=np.max), y_axis='log', x_axis='time')
plt.title('Power spectrogram')
plt.colorbar(format='%+2.0f dB')
plt.tight_layout()
plt.show()
```

MFCC の算出・描画プログラム

```
import wave
import numpy as np
import matplotlib.pyplot as plt
import librosa.display
import librosa.feature

music, fs = librosa.audio.load("cuckoo1.wav")

stft = np.abs(librosa.stft(music, n_fft=2048,
hop_length=512))**2
log_stft = librosa.power_to_db(stft)

S = log_stft

# メルフィルタバンク
mel_basis = librosa.filters.mel(sr = fs, n_fft = 2048,
fmin = 2000)
melsp = np.dot(mel_basis, S)

mfccs = librosa.feature.mfcc(S=melsp, n_mfcc=13)

# 標準化して可視化
import sklearn
mfccs = sklearn.preprocessing.scale(mfccs, axis=1)
librosa.display.specshow(mfccs, sr=fs, x_axis='time')
plt.title('Mel frequency Cepstrum Coefficients(0Hz)')
plt.colorbar()
plt.tight_layout()
plt.show()

#平均化
result = np.zeros((mfccs.shape[0],1))
result[:,0] = np.mean(mfccs,axis=1)

print(result.flatten())
```

