

東京大学大学院新領域創成科学研究科
社会文化環境学専攻

2016 年度
修 士 論 文

野生動物装着型センサの位置推定手法のための周辺環境音
のクラスタリングに関する研究
Location Estimation for Animal Wearable Sensors by using
Peripheral Environmental Sound Classification

2016 年 1 月 23 日提出
指導教員 瀬崎 薫 教授

合 間 優 陽
Yuya, Kamma

目次

第1章 序論	1
1.1 はじめに.....	1
1.2 本論文の構成.....	2
1.3 研究背景.....	2
1.3.1 実地調査に伴う課題.....	2
1.3.2 既存手法と問題点.....	5
1.3.3 物理的な立ち入りを禁じられた区域.....	6
1.3.4 野生のニホンザルを用いたセンシング.....	8
1.3.5 動物間通信 (CPSS) を用いたセンシング.....	9
1.3.6 CPSS の問題点解決・本研究の目的.....	10
第2章 関連研究	12
2.1 航空音の騒音を機械学習により判別する手法の提案.....	12
2.2 電気・ハイブリッド自動車の近接検知.....	12
2.3 野球試合の実況者音声を用いた場面転換時の抽出.....	14
2.4 音声認識手法を用いた嚙下音による水分摂取量推定手法.....	15
第3章 航空音の利用により位置同定をおこなうまでの提案システム	16
3.1 提案システム —航空音認識から位置同定へ—.....	16
3.2 音声認識と機械学習を用いた航空音の自動分類への試み.....	19
3.2.1 サポートベクトルマシン (SVM).....	19
3.2.2 メル周波数ケプストラム係数 (MFCC).....	21
3.3 航空機ジェット音をはじめとした環境音の種々たる特徴.....	22
第4章 提案システムの評価実験	27
4.1 実験の概要 —航空音識別手法—.....	27
4.2 実験システム・条件.....	28
4.3 実験結果.....	31

4.4	実験評価	33
4.5	本節における実験での問題点	36
第5章	ローパスフィルタによる提案システムの改善	37
5.1	提案システム改善案の概要 —フィルタ通した識別—	37
5.2	改善案を用いた実験結果	39
5.3	改善案の評価	39
第6章	結論	43
6.1	本研究のまとめ	43
6.2	本研究の残された技術的な課題	44
6.3	今後の可能性	46
	謝辞	49
	発表文献	51
	参考文献	52

目次

図 1.1	日本国土における通信カバーエリアマップ	3
図 1.2	センサ毎における観測画像および分解能の違い	5
図 1.3	福島県における避難指示区域の状況概念図（2016年10月28日現在）	7
図 1.4	DTN を用いた動物間マルチホップ転送システム（CPSS）の概念図	9
図 2.1	環境音毎におけるスペクトラムおよび航空音の特徴	13
図 2.2	EV および HV 車におけるスイッチングノイズ	13
図 3.1	提案システムの概念図	17
図 3.2	測位手法の概念図	18
図 3.3	マージン最大化の概念図	20
図 3.4	ケプストラム抽出の流れ	21
図 3.5	メルフィルタバンク作成の過程	22
図 3.6	環境音測定所イメージ	23
図 3.7	船田における航空音（上）および野鳥音（下）のスペクトラム	24
図 3.8	栃本における航空音（左）および野鳥音（右）のスペクトラム	25
図 3.9	矢竹における航空音（左）および野鳥音（右）のスペクトラム	25
図 4.1	単純処理における特徴量	28
図 4.2	航空音自動認識システムのフローチャート図	29
図 4.3	分類結果についての精度を表す混同行列	31
図 5.1	ローパスフィルタによる野鳥音の除去	38
図 6.1	野生動物にセンシングデバイスを搭載するにあたる技術的難易度（左）	45
図 6.2	全世界における自然保護区域（右）	47
図 6.3	アオウミガメを用いた海洋環境の空間情報センシング機構の概要図	47

表目次

表 3.1	各音源の記録形式など詳細.....	26
表 4.1	単純処理におけるカーネル毎の分類結果.....	32
表 4.2	MFCC 処理におけるカーネル毎の分類結果.....	32
表 4.3	単純処理における分類正解率.....	33
表 4.4	MFCC 処理における分類正解率.....	33
表 4.5	ROC 曲線の成り立ちをイメージするためのサンプルデータ.....	33
表 4.6	単純処理におけるカーネル毎の ROC 曲線.....	34
表 4.7	単純処理での ROC 曲線における AUC 値（航空機と全体平均のみ抜粋）..	34
表 4.8	MFCC 処理におけるカーネル毎の ROC 曲線.....	35
表 4.9	MFCC 処理での ROC 曲線における AUC 値（航空機と全体平均のみ抜粋）	35
表 5.1	単純および MFCC 処理における 3 および 4 クラス毎の分類結果.....	38
表 5.2	単純および MFCC 処理における分類正解率.....	38
表 5.3	3 クラスにおける双方処理の ROC 曲線.....	40
表 5.4	3 クラスにおける双方処理での AUC 値.....	40
表 5.5	4 クラスにおける双方処理の ROC 曲線.....	41
表 5.6	4 クラスにおける双方処理での AUC 値.....	41

第 1 章

序論

1.1 はじめに

2011 年 3 月 11 日に東日本大震災が起きてから早くも 6 年が経つが、この震災により福島第一原子力発電所の事故が起きたことは記憶に新しい。この事故により、福島第一原子力発電所から半径 20km、または、非常に高いレベルの放射線量が検出されている地域は 2017 年の今日に至っても、未だに立ち入りが禁止されている [1]。そのような立ち入りが禁止されているような地帯において当時飼育されていた動物が置き去りとなり野生化していることが知られている。そのみならず、原発から放射性物質が漏れだしたことにより周辺の森林や生態系への影響は甚大なるものになるであろうと懸念されている。そのような観点から、この立ち入り禁止の地域において環境および生態系がどれほどの影響を受けているのかを早急に調査を行うことが必要とされている。

しかし、立ち入り禁止の区域ということもあり継続的に研究員がその地域において調査を行うことは、極めて高い線量を浴び続けることや現地調査を行う際に設営する設備が与える環境への配慮が必要とされる。この他にも様々な問題が考えられ、なおかつその問題における対策は時間と労力と時間をかけることになるであろう。

上記のことから、現地へ調査員を派遣し環境調査を行うことは様々な問題が生じると考えられる。しかし、近年、モバイルデバイスを始めとした情報通信機器および半導体の小型化など技術の発展がめざましく、それらを用いることにより調査員が継続的に現地に向かうことなく間欠的にフィールド調査を進めることを実現するための研究が行われている [2] [3] [4] [5]。これらの研究により、調査員に依った現地のセンシングデータを無人で取得することが可能になることで現地調査におけるリスクが軽減されることが考えられる。また、間欠的な運用を行うことにより、より密な調査結果を得ることで環境や生態系の時系列的な変化を捉えることができよう。しかし、このような手法は現在も研究中の内容であり、まだ確立されていないのが現状である。

本研究はその手法および研究の一部として貢献する内容となる。具体的には、後述する動物間通信を用いた環境センシングの際に動物の位置情報を把握する必要がある。その際

に、上空を飛行している航空機の音を用いることによりおおよその位置を推定する。すでに、この推定手法は確立されているものの、調査員が手計算により算出しているのが現状である。そのことから今後、手計算から全自動化を行う必要がある。この推定の際には、航空機の音を用いるのであるが、現地調査の場合であれば実際に音を聴くことにより手動でラベル付けなどが可能である。しかし、既存の研究では機械に航空音を判断させる手法がまだ確立されてなく自動化するうえで必要な技術とされる技術となる。本研究では、その自動化をするにあたる環境音のクラスタリングに関する手法の提案から検証を行ってゆく。

1.2 本論文の構成

本論文は、以下のような構成で成り立つ。また、章ごとにおける概要を述べる。

第 1 章「序論」では本研究をおこなうにあたる背景を述べる。

第 2 章「関連研究」では関連する研究分野について挙げる。また、本研究における可用性や問題点について示す

第 3 章「航空音の利用により位置同定を行うまでの提案システム」では本研究における提案システムについて記す。その際に、システム上で用いる技術の説明と扱う特徴量について述べる。

第 4 章「提案システムの評価実験」では第 3 章において提案した手法の実験と評価について示す。

第 5 章「ローパスフィルタによる提案システムの改善」では 4 章の実験において考えられる問題点をあげ提案システムの改善を図った内容について述べる。

第 6 章「結論」では 4 章と 5 章により得られた結果についてまとめる。また、本研究における技術的な課題から今後の展望について記す。

1.3 研究背景

本節では研究背景として、電源や情報インフラなどの利用が限られた地帯における環境調査の問題点や課題をとり挙げ、さらに、そこで調査を行う際に生じうる問題を解決へと導くために行われている試みなどを述べてゆく。

1.3.1 実地調査に伴う課題

情報通信端末を用いた環境調査（モバイルセンシング）がスマートフォンや情報通信網、そしてそれらを用いたセンサネットワークの技術的な発展もあり近年盛んに行われている。

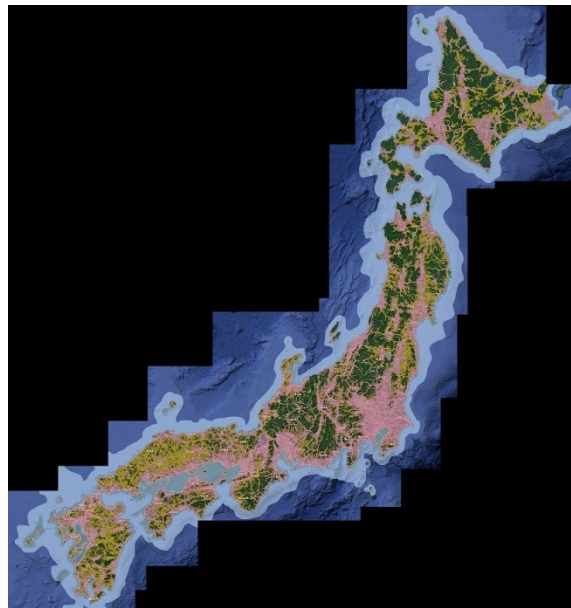


図 1.1 日本国土における通信カバーエリアマップ¹
(緑色の箇所が未カバーエリア)

これらの恩恵もあり、モバイルセンシングがオフィスや繁華街がひしめく環境を想定した都市部や人間が生活を営む環境を想定した郊外であれば個人レベルでも行うことが可能となった。しかし、これは安定した電源や情報インフラが存在するからこそ行えることであり、野生動物などが存在するような森林や山間部などの自然環境においては、限られたインフラのみしか用いることができない。これより、上記のような都市部や郊外におけるセンシングは難しいのが実情である。この経緯から、特殊な場所における調査は未だに現地に調査員を派遣し、様々な情報やデータを収集しに行く手法がとられている。

現地に調査員が直接向かうからこそ得られる情報はあるかもしれないが、自然環境、特に人により手を付けられていないような場所に人が立ち入ることで、その環境が現在進行形で破壊されてゆくということが考えられる。また、持続的かつ間欠的な調査を行うとなれば破壊は進み、現地に何度も赴くことは環境保全のみならず設備維持などのコストにおいても効率的ではないと思われる。さらに、そのような自然環境が必ずしも人間にとって安全な地帯であるとは限らない。例えば、1.1 節で挙げたような環境である。以降ではこれらの課題に関して述べてゆく。

森林や山間部における自然環境では限られたインフラしか使用できないと先述したが、日本の国土の約 70%は森林であり [6] その多くの地域は通信圏外となる。これは、山間部や森林の多くの地域が未だ通信がカバーされていないことが図 1.1 からわかる。これは、モバイルセンシングを行うことを困難にすることのみならず、実地調査を行う際における

¹ 図は NTT ドコモのエリアマップ [33]を基に画像加工を施し作成

想定外の事態（例えば、遭難や怪我などの救助要請）が発生した際に、専用の機材を携帯していない場合は外部との連絡をとることができないリスクが伴うことがあり得る。また、日本生態学会は山や陸上における現地調査での事故を以下のように類型化している [7]。

- 気象遭難（吹雪・大雨・台風・雷）
- ルートミス
- 突発的災害（落石・雪崩・土石流・鉄砲水・氷河崩壊・火砕流・山火事・倒木 など）
- 突発的事故（滑落・転落・雪庇踏み抜き・雪洞崩壊・樹木からの落下・テントの火事）
- 衰弱・低体温症（疲労凍死）
- 疾病・ケガ（高山病・心臓疾患・熱傷・凍傷・骨折・熱射病・熱疲労・感染症 など）
- 危険生物（ヒグマ・ハブ・スズメバチ など）

実際に、警視庁生活安全局地域課が発行している平成 27 年における山岳遭難の概況 [8] によると道迷いが全体の約 4 割、滑落・転倒が約 3 割を占めている。道迷いが多い理由としては、やはり山間部において携帯電話が使用できないことが起因していると考えられる。しかし、このデータにおいては、一般の登山客などを対象にしている。これより、専門の調査員であればリスク管理や研修などが行われているであろうことから、事故発生リスクは低減されるのではないかと考えられる。しかし、全国地質調査業協会連合会が発行した地質調査業務における労働災害の現状と対処 [9] によれば実地調査中の事故は、ボーリング作業中の事故が全体の約 4 割、転落・落石による事故は約 2 割を占める。これより、地質調査を行うような専門の作業員であろうとも災害に遭う可能性があると言える。また、ボーリング作業に伴う事故は地質調査ならではと言えるであろうが、転落・転石における事故は他の実地調査においても共通して想定できる事故であると言える。

実地調査における事故は数値のみならず、事例が過去に報告されている。例えば、山林における用地測量中に技術員が滑落死したこと [10] や毎木調査の際にスズメバチに臀部を刺された [11] と言った事例が挙げられる。用地測量に伴う作業中に起きた事故に関しては、林木調査においてその範囲を決めるためにビニールテープを張る作業中に生じた。その時の作業員は 2 名であり、作業を終え各々手荷物置き場へ移動する際に片方の作業員が山林を滑落し、死亡したと記されている。事故の発生原因としては、作業手順の不備や安全帯を不適切に使用していたといった理由が挙げられている。これは、上述した類型化において、突発的事故に分類されることがわかる。また、毎木調査の際に生じた災害では、選木作業中に蜂の巣を刺激したことから作業員が臀部をスズメバチに刺された事例が報告されている。これは、類型化において危険生物の括りとなる。滑落などの事故は本人が細心の注意を払うことにより生じるリスクは減らせるであろうが、危険生物に関しては防ぎようが殆ど無いのが事実である。

これらの事故から共通して言えることは、その分野に長けた人物であろうとも不慮の事

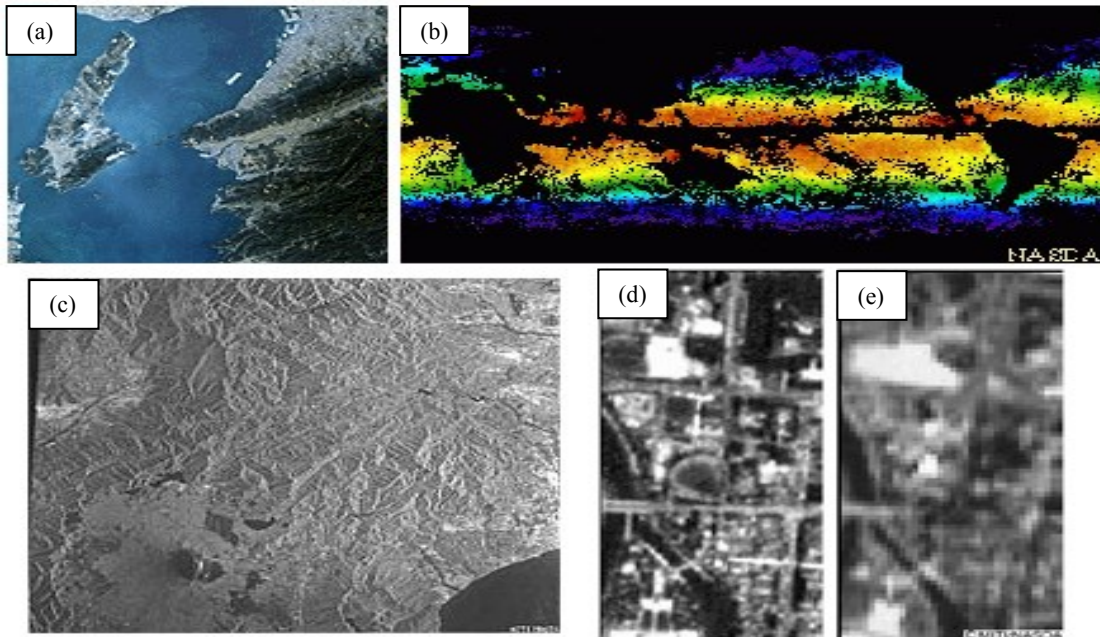


図 1.2 センサ毎における観測画像および分解能の違い²

故に遭う可能性があると言えよう。それゆえ、実際に人が実地調査を行うことはある程度
 のリスクを抱えることが分かる。そして、それが山間部や立ち入り禁止区域のみならず極
 地や熱帯林のような非日常的な地帯になればなるほど、よりリスクの高いものになるであ
 ることが想定できる。現地に赴き真値を得ることで情報の有用性や可用性を見出すこと
 は、森林や低山などであれば通用することであり、非日常的な地帯となるとそう簡単には
 いかないであろう。さらに、そこに滞在し続けること自体が難しくコストが生じることに
 なる。

では、このようなリスクを減らすためにはどのようなことができるだろうか。それは、
 人間が極力観測地へ向かわなければ良いのではないだろうか。現地へ向かわなければ、安
 全面に関しては確保される。また、長期的な設備の運用や滞在を行わなくて済むことなど
 からコスト面での解決が期待できる。何よりも、手付かずの自然において人が立ち入らな
 いことにより、その状態が維持されることは大きなメリットとなるであろう。

1.3.2 既存手法と問題点

1.3.1 項においてとりあげた問題点を解消したような手法が存在する。それは、リモート
 センシングと呼ばれる。リモートセンシングとは、光学センサやマイクロ波センサを搭載

² 図は宇宙航空開発機構の Sensor types および Resolution of Sensor [34]より
 なお、(d)の衛星名は ADEOS, (e)の衛星名は JERS-1 であり Landsat-8 とは異なる

した地球観測衛星を利用することにより対象地域を観測する手法である。具体的には、衛星により撮影された地球外からの地形画像（Landsat 画像と呼ぶ）が宇宙から衛星観測所に送信される。その送信された Landsat 画像には搭載されたセンサに応じてデータが付加されている。その撮影された地域における画像に付加されたデータを解析することで年月経過による森林や農耕地帯の時間的差分などを比較することが可能である。その他にも、地形の長期間における変化度合いや、大陸における気温変化の調査などを可能とする。この画像はインターネット通じて一般に公開されていることから、ArcGIS や QGIS などの地理情報システムソフトを用いることにより誰もが環境センシングを行うことを可能としている。また、この手法を用いることにより極地における観測を現地に人を派遣することなくセンシングを行った研究も発表されている [12]。このことから、既存手法により現地へ調査員を派遣せずにセンシングを行うことは、可能であると言えよう。

しかし、当手法には問題点がいくつかある。ひとつは、観測衛星に搭載されたセンサの性能によって取得できるデータに大きな差が生じてしまうことである。図 1.2 の(a)は可視光を、図 1.2 の(b)は熱赤外線、図 1-2 の(c)はマイクロ波を用いたセンシング結果となる。それぞれの観測方法には特徴がある。可視光を用いた手法は分解能が比較的高いが太陽光を用いた観測になるため雲の下や夜間の観測ができない。熱赤外線を用いた手法は、地表や海面の表面温度を計測することができる。マイクロ波センサを用いた手法は、本来雲が画像中にあったとしてもその下を観測することができるメリットがある。しかし、分解能が比較的低いことが挙げられる。これは、図 1-2 の(d)と(e)を比較することでも分かる。(d)は光学センサを用いた衛星画像であり、(e)はマイクロ波センサを用いたものである。

一般に提供されている Landsat-8 号における衛星画像の分解能（地上の物体をどの程度の範囲まで見分けられるかの尺度）に関しては可視光線のデータでは 15~30m、熱赤外線のデータでは 100m と言った分解能である。これは、広範囲のセンシングであれば問題はないであろうが、狭い範囲においてはより高い分解能が必要とされる。他の問題点としては、観測衛星は 16 日周期で地球上を周回していることから連日のデータを取得できないことや、衛星画像は非常に大きなデータ量であることもあり、地球のデータセンタへ送信され即時公開されたとしても利用できるまでに 2 時間ほどの時間を有する [13]。

これより、リモートセンシングは自然環境を破壊することなく、また、調査員を現地に派遣することを必要としないような環境調査を行えることがこの手法の強みである。しかし、連続性や即時性を求められるような環境センシングにおいては不向きであることが言えよう。

1.3.3 物理的な立ち入りを禁じられた区域

立ち入りが禁じられる地域とは、原生林などの自然環境を保全することや人間にとって生命を脅かすような危険地帯のいずれかが考えられる。生命を脅かすとは、例えば武装地

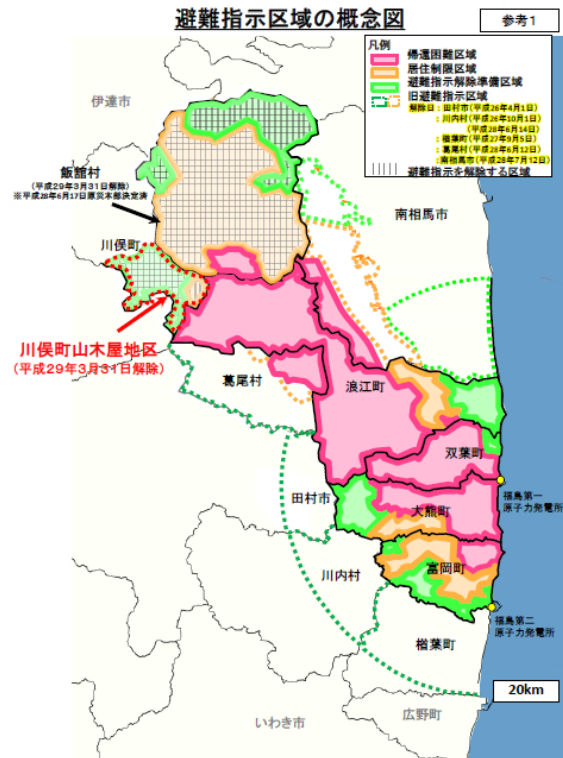


図 1.3 福島県における避難指示区域の状況概念図（2016年10月28日現在）³

帯や天然有毒ガス発生地帯，そして原発事故跡地などが考えられよう．ここでは，原発事故跡地について着目してゆく．

これまでに，国際原子力機関（IAEA; international atomic energy agency）が定める国際原子力事象評価尺度（INES; international nuclear events scale）による最高レベル7を認定された原発事故は2つある．1つはチェルノブイリ原子力発電所事故，そしてもう1つは福島第一原子力発電所事故である [14]．レベル7がどれほどの事故指標なのかとして INES はこう定めている．

“Major release of radioactive material with widespread health and environmental effects requiring implementation of planned and extended countermeasures.”

－（IAEA が定める INES 本文より引用 [15]）

この指標に鑑みれば，この2件の原発事故は炉心内の放射性物質を大量に外部放出し，人間のみならず広範囲の自然環境に影響を与えた事故と言えよう．実際に，チェルノブイリの原発事故により大気中に放出された放射性物質は，ヨウ素換算で520万テラベクレル，

³ 概念図はふくしま復興ステーションから引用 [1]

福島原発事故では 77 万テラベクレル、あるいは 57 万テラベクレルと推定されている [14]。人間はある一定の放射線量を被曝し続けることにより死亡、および重度な後遺症を残すこともあることから、2017 年現在、福島原発近辺では未だに立ち入り禁止とされる区域が設定されている [1]。図 1.3 はその区域について表した概念図である。赤色の線により囲われた区域については、帰還困難区域とされ部外者はおろか住民でさえもその区域において常時立ち入りは許可されていない。仮にこのような区域においてセンシングを行うとなれば、多大なるリスクやコストが伴うと想定される。そもそも立ち入りを許可してもらえないかどうかも分からない。たとえ立ち入りを認可されたとしても、連日・長時間の調査を行うことは困難であることが想定される。

福島第一原子力発電所事故の際には、原子炉の調査において早急なる事故発生原因の究明が要された。しかし、原発は建物内に存在していることより屋内調査であったことや、連続的なデータを取得できないと言った背景から前節で挙げたリモートセンシングによる調査は不適であった。そこで、この時はロボットを用いた手法で実地調査を行った実績がある [16]。人間が遠隔操作によりロボットを扱うことで、原発内の線量をセンシングすることにより、どこから放射線が漏れ出しているのかを調べるといった調査に役立った。しかし、この手法にも問題点があり、センシング先でロボットが故障を起こしてしまった場合に対処のしようが無いことや、製作において多額なコストがかかることが挙げられる。実際に調査の際にロボットが故障を起こし、一時的に調査が中断したと言う内容が報道されている。故障の原因としては、震災の際に生じた障害物の多さを一因として挙げている。このような人間が建設した比較的障害物が少ない状態であることが予期される建物や施設内であれば、ロボットが巡航するにおいては問題がないであろう。しかし、山林など起伏や障害物の多い場所においては適さないことも考えられる。また、ロボット型の調査機を山林などで用いるとなると、動物は人間と比較し耳が良く、そのような機械音に敏感であることや縄張りに侵入されることでストレスを感じ、その地から別の地域へと姿を消すかもしれない可能性が考えられ本来の手付かずの状態であった自然の形が壊れる原因となることが懸念される。このことから、ロボットを用いた環境調査は屋内の平地である場合に対して有効であると言えよう。

1.3.4 野生のニホンザルを用いたセンシング

1.3.3 項で述べたような背景から、福島大学の高橋ら (2011) は、これまでに取り上げてきた福島第一原子力発電所事故の影響による森林の汚染の状況を把握するために、野生のニホンザルを用いた手法を考案した。具体的には、ニホンザルに線量計および全地球測位システム (GPS) を搭載した首輪を装着させることにより実験を行った。実験の流れは、サルに首輪を装着し山林に放し放射線量と位置情報を記録させ、サルの首から外れた測定器を回収し記録を分析するものである。なお、首輪は特定の信号を測定器に送信する

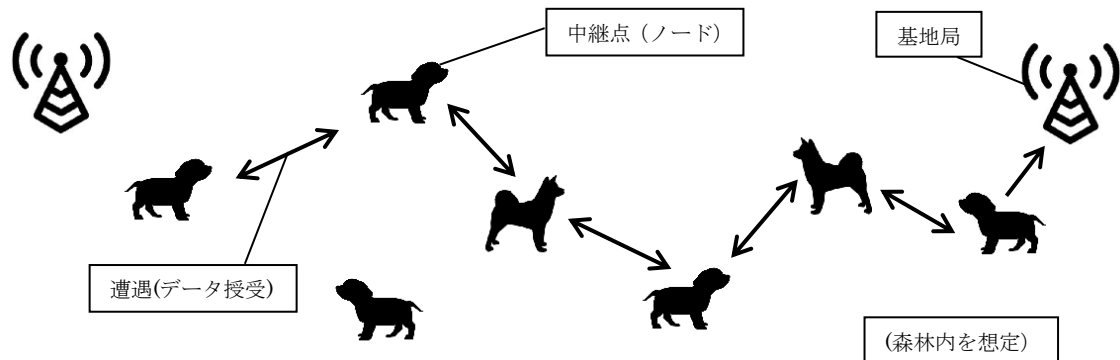


図 1.4 DTN を用いた動物間マルチホップ転送システム (CPSS) の概念図⁴

ことによりサルの首から外れるようになっているとのことである。同研究グループは、群れからはぐれにくい雄ザルを捕獲し装着することによって一定エリアの放射線量の分布を把握できると考えた。しかし、サルの生息地は限られるため、観測の範囲が限られてしまう問題点が生じた。何よりも、線量の記録を分析するためには首輪型センサを回収しに行く必要があり、結局は人間が実地に立ち入る必要がある問題が挙げられる。

1.3.5 動物間通信 (CPSS) を用いたセンシング

これまでに挙げた実地センシングにおける問題点を踏まえ、野生動物が生息して電源や情報インフラ網が極めて限られた地域において人間がデバイスなどを回収すること無くデータを取得できる手法として、動物を通信の中継点として最終的に基地局までデータを搬送させる手法が考えられた。

中川らは、遅延耐性ネットワーク (DTN; delay tolerant networking) 技術を用いた動物間マルチホップ転送システム (CPSS ; carrier pigeon-like sensing system) を提案した [5]。DTN 技術とは、ある地点から情報を任意の目的地まで送信する際に、通信環境が劣悪な地域においても多数の中継点 (ノード) を設定することにより目的地までデータを転送するという技術である。図 1.4 に CPSS の概念図を示す。この技術では、Arduino により作製されたデバイスを搭載した動物 (ノード) 同士がすれ違った際に Bluetooth や ZigBee などの近距離無線通信技術を用いることにより自身が所持している情報を他のノードへ伝達していく。このリレーを繰り返すことにより、最終的には外部にデータの送信できるような基地局にまで動物により計測されたデータが搬送される仕組みとなる。また、DTN は通信環境が劣悪な地域においても通信環境を構築する技術であることが想定されている [17] ことから電源や情報インフラ網が限られたような自然環境において適した技術であること

⁴ 中川ら [5]を参考に作成

言えよう。

しかし、動物同士でデバイスを介した通信を森林内で行うにあたり、デバイスには電池が内蔵されている。しかし、周辺にノードである動物が存在しないような状態にもかかわらず、常時データの送信を続けていることは電池切れを起し、データの記録を行えなくする要因となることが懸念される。そこで、中川らは動物の習性を利用し、動物がすれ違った際のみデータを受受するような手法の構築を行った。具体的には、例えば犬の場合であると、犬は犬同士ですれ違う際にお互いの臭いを確かめる習性があるとされる。その際に加速度センサにより得られた値が、予め定義したすれ違いを行っている判断される閾値を越えていた場合に、すれ違いを行ったと判断しデータの送受信を行うといったものである。また、牧山らはこのすれ違いを検出する際のアルゴリズムの検証や人間の行動検知を動物に応用した検出アルゴリズムの提案を行っている [18]。その検証結果より、中川らが提案したアルゴリズムは有効であることが言え、また提案された検出アルゴリズムによりさらなる検出精度が向上されたことが報告されている。ただし、動物ごとによってそのアルゴリズムが通用するのかどうかについては、今後も検証を行う必要性があると述べている。

この手法を用いることにより、立ち入りが制限されたような区域においても野生動物を用いることによりリモートセンシングにおける連続的なデータの取得に関する問題点は解決できるように思われる。しかし、この手法にも問題点はあり、一つは位置情報を記録できない点である。これは、電池消費量を減らすことや山間部においては衛星測位が難しいと言ったことが挙げられる。また、動物に積載できる物体の重量は動物における体重の 2% 程度と制限がある [19] ことが挙げられることより、重量のあるセンサやデバイスを搭載することが難しいことが挙げられる。さらに、搭載したデバイスはインターネットに接続できないこともあり時刻同期ができない点が挙げられる。

1.3.6 CPSS の問題点解決・本研究の目的

1.3.5 項において述べた背景より、この技術における問題点の解決に向けて、我々はいくつかの手法を考えた。立ち入りが制限されたような区域かつ森林などにおいては利用できるリソースは少なく、そのような区域において常時利用できるようなリソースとして、我々は以下の 3 点を挙げた。

- 地鳴り（地響き）
- 防災無線を用いた時報
- 航空音

最初に、地鳴りについてであるが、これは発生の際におけるエネルギーが大きいことや発

生時刻などの詳細な情報が得られることから、有効なリソースであるように感じられる。しかし、発生するタイミングが予想できないことや、その発生間隔が不定であることから、今回用いるリソースとしては不適であると見込んだ。次に、防災無線を用いた時報であるが、これは定期的に音源を得ることができ、時刻同期を行う際には都合が良いように感じられる。しかし、大きな問題点として、夜の時間には時報を放送しない自治体が多数である。このことから、時報を用いることは断念した。最後に、航空音であるが、航空音に関しては24時間上空を飛んでいることや航路が定まっていることから観測地帯周辺においての飛行が望める場所であれば安定したリソースとして用いることができると考えられる。また、インターネット上で航空機の位置情報から高度などの情報を提供するサービスが存在する [20]。このことから、我々は航空音を用いることがこの問題に対して有効な情報源となるのではないのかと言う結論に至った。

具体的な手法は、図 1.4 のようなエリアの基地局に固定マイク、動物に積載した端末に小型マイクを搭載することで航空音を記録する。センシング地域が広域になれば固定マイクに記録された航空音の記録時刻に、音の到達時刻の観点から、差が生じると想定される。その差を利用することにより、動物のいた場所の推定を行う。しかし、この手法を試みるためには航空音がいつ鳴り響いたのかと言うことを検知できるようにしなければならない。航空音が鳴るたびに手動で時刻や位置などをラベル付けするのであれば、結局は現地に調査員を派遣することとなり当初と何ら変わりはないことになるからである。

このような事由から、本研究では、CPSS に存在する問題点を解決するための一端として貢献することが本研究の目的となる。

また、本研究の目的を達成すべく、以下のような流れで実験および評価を行った。

- 環境音に音声認識処理をかけることが環境音分類において有効な処理であるのかどうかの実験および評価
- 環境音分類を行う際にさらなる精度向上を図る手法の提案、およびその実験内容の評価

以降では、環境音より航空音を自動認識・分類するための実験から評価までについて記してゆく。

第2章

関連研究

計算機に音声を自動で識別させるには様々な手法が考えられるが、近年は環境音に音声処理を施し機械学習にかけることで自動識別を行う研究が行われている。本章では、環境音を自動認識するためにこれまで行われてきた研究および手法を紹介することから、本研究における可用性や問題点について示す。

2.1 航空音の騒音を機械学習により判別する手法の提案

Bertrand [21]らは、航空音によりもたらされる騒音の自動判別手法を提案している。具体的には、航空音を機械学習と音声処理を用いることにより自動で判別を行う方法である。しかし、この論文は自動判別が主としたテーマではなく、空港周辺における騒音問題に関する論文であったこともあり、手法に関しては提案のみに留まっている。つまり、実際にデータを収集した実験や評価は行っていない。

しかし、図 2.1 でもわかるよう、様々な乗り物における音声の特徴を挙げ、さらに航空機におけるスペクトラムでは 2kHz 帯に特徴的な傾向が伺えることを示していた。また、4kHz 以降では音圧が低まる傾向を示している。この論文内では 2kHz 帯と、高周波域において特徴的な音を出しているかどうかに関しては検討する必要性があると言える。なぜなら、高周波域において特徴的なスペクトラムを出していれば分類精度が向上する可能性が高まるからである。その理由については、次節において述べる。

この研究より環境音を自動認識するには、音声を機械学習の分類器にかけることにより判別を行うことが手法の一つであることが伺える。そこで、航空音ではない音源を分類した手法について他の研究例を示す。

2.2 電気・ハイブリッド自動車の近接検知

高木らはスマートフォンのマイクにより録音された音源を用いた、電気自動車 (EV 車) と

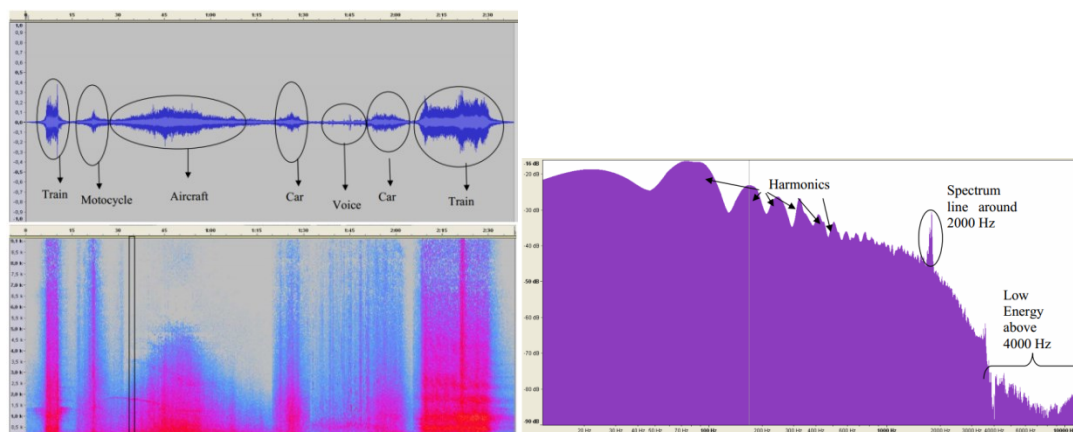


図 2.1 環境音毎におけるスペクトラムおよび航空音の特徴⁵

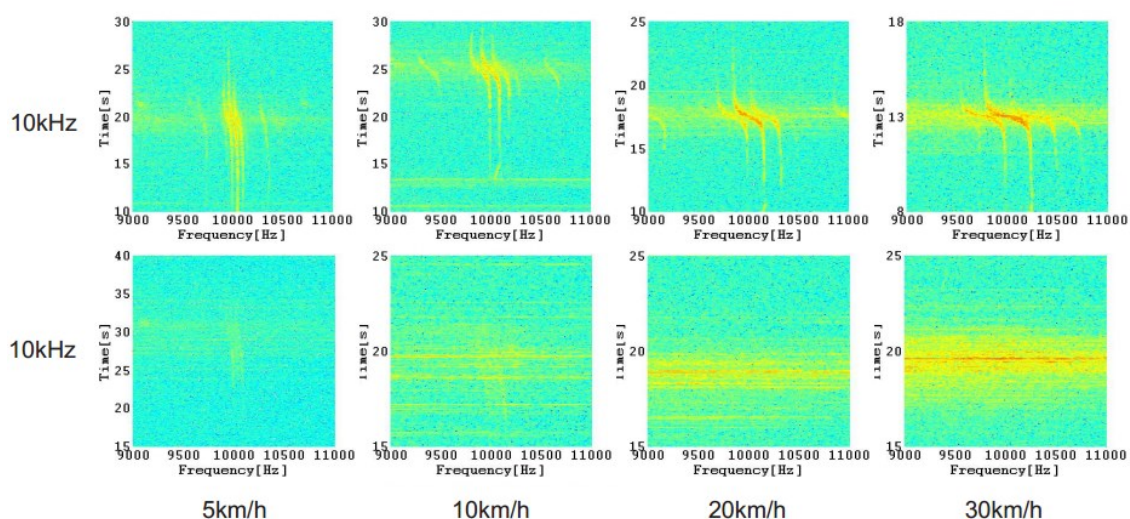


図 2.2 EV および HV 車におけるスイッチングノイズ⁶

(上段は EV, 下段は HV 車)

ハイブリッド車 (HV 車) の分類に成功している [22] . この研究では EV および HV 車から発せられる特徴的な周波数 (スイッチングノイズと言い, 高周波帯に生じるノイズ) に着目することで分類を実現している.

分類は, 実際に録音した自動車走行音に車種 (EV か HV 車) や車速 (5, 10, 20, 30Km/h), そして車両の有無をラベル付けした音源を高速フーリエ変換 (FFT; fast fourier transform) にかけて後, 分類器にかけることにより行う.

この研究内では機械学習としては教師あり学習で分類器は WEKA (J48 決定木) を用いている. また, 検知精度の評価に関しては, 10 分割交差検証を用いることにより行っている.

⁵ 図は Bertrand [21] を基に画像加工を施し作成

⁶ 図は高木ら [22] を基に画像加工を施し作成

る。実験結果では、EV 車は 5km/h 程度の低速であれば 97.06%と非常に高い精度、30km/h では 76.47%。HV 車では、5km/h で 70.51%、30km/h で 50.0%の精度であることが判明した。論文内では、この結果から、EV と HV 車は分類器を用いることで判別することが可能であると結論付けた。

この結果から、この場合であると EV や HV 車のような特定の環境音においては分類器を用いた機械学習による音源識別は可能であると言えよう。しかし、今回の例であるとスイッチングノイズのような高周波帯に特徴的なスペクトラムが発生するような音源でないと識別できない可能性が考えられる。実際に、HV 車の 30km/h では 50.0%とランダムで分類を行った精度とほぼ変わらない結果が出ている。この原因として高木らは、図 2.2 を見て分かるように、HV 車におけるスイッチングノイズは高速度になるにつれてスイッチングノイズがほとんど見えないことに起因していると述べている。また、録音した音源は FFT をかけ全周波数帯域での平均音量で正規化を行ったものであり、他の処理を行っていない。

2.3 野球試合の実況者音声を用いた場面転換時の抽出

Rui らはテレビ放映された野球の音源を用い、自動的にハイライトを抽出する機能を提案している [23]。この機能を提案した背景としては、海外におけるテレビ番組のチャンネル数が増加し、視聴時間を削減すると言った目的が挙げられている。また、野球を対象とした理由としては試合の起伏(実況者の声のトーンや観客の声援の大小など)が大きいこと、それに加えて試合時間が比較的長いと言う点から、ハイライト抽出において野球が適しているとしている。

野球放送では試合の展開に影響を与えそうな環境音として、解説者の実況音やヒット音などと試合固有の音や観客の声援や拍手などが考えられる。しかし、ここでは試合展開とは関係がない箇所でも起こり得る観客の声援や拍手などは特徴点として見做さないことにしている。また、試合が動く瞬間とはピッチングを行いそしてヒット生じた後に発生するものであり、それはアナウンサーのスピーチに相関するものであると仮定している。そのような過程から、ハイライト抽出をする際には野球実況者の音声を特徴量として用いている。

この場面展開の抽出を行う際に、実況者の音声を用いているが具体的には以下の 3 つを試合の起伏を抽出する際の特徴量として挙げている。

- 入力音声のエネルギー (蝸牛フィルタに基づき分割を行った値 [24])
- エネルギーと入力音声における平均情報量 (エントロピー)
- エネルギーと MFCC (MFCC については後述)

また、場面の展開は、上述した特徴量とヒット検出の確率を足した融合格率と言う尺度を

用いることにより算出している。これらの特徴量で実況者が興奮、つまり試合が動いていると思われる音声を機械学習により予測を行う。

この条件により分類を行ったところ、人間がラベル付けした試合が動いているかどうかと言う場面のデータに対して 75%の精度で正しい分類が行えた結論付けている。なお、その際に用いる機械学習における分類器についてであるが、野球のような試合における音源では、実況者の音声のみならず観客の声援や楽器の音源が混合したノイズに対して堅牢性が求められる。そこで、Ruiらは学習手法としてサポートベクトルマシン (SVM; support vector machine) や、ガウシアンフィッティング (GAU; gaussian fitting) そして、k近傍法 (KNN; k-nearest neighbor algorithm) の3種類の学習法で性能を比較した。その結果、3つの学習方法のなかで SVM が最も精度が高い結果を得ることができたことを示している。

2.4 音声認識手法を用いた嚙下音による水分摂取量推定手法

山田ら [25] は環境音として人が水を飲む際に発する嚙下音を用いることにより、水分摂取量を推定する手法を提案している。この提案手法では、嚙下音の特徴量を抽出する際に不特定の被験者に対し、共通かつ適切な特徴量を算出できる音声認識手法としてメル周波数ケプストラム係数 (MFCC; mel-frequency cepstrum coefficients) を用いている。MFCC は従来、人の発話内容識別に有効であるとされた手法であるが、この研究では人の発話以外における音源において用いている。実際に、従来の手法と比較し、被験者における水分摂取量推定の誤差を減らすことに成功し、精度向上を実現した。このことから、MFCC が人の音声以外の音響認識にも有効な手法であると考えられる。

また、行動分類 (深呼吸, 咀嚼, 咳, 発話, 嚙下の 5 種類) を行う際には特徴量として Relative-MSFS (mel-scale fourier spectrum), 分類器は隠れマルコフモデル (HMM; hidden markov model) を用いた実験も行っている。その際には評価指標として、適合性, 再現性, F 値を用いることにより分類結果を評価している。

この実験の結果より、MFCC を音声以外の内容に対しても特徴量として用いることが有効であることが伺える。また、単純な音圧を特徴量として用いることとの比較手段としても有効であると考えられる。

第3章

航空音の利用により 位置同定を行うまでの提案システム

本研究では、1.3.6 項においてあげた航空機のジェット音から野生動物の位置を同定するシステム、およびそれに伴う全自動化へ向けた手法を提案する。最初に、システムに関しての説明を行い、次に自動化を行う際に必要となると思われる技術の説明を行い、その次にシステムの構成を説明する。

3.1 提案システム —航空音認識から位置同定へ—

これまでも挙げたよう、立ち入り禁止区域において使用できるリソースは限られている。そのような環境でも、ある程度の短いスパンで用いられる情報源として、航空音が挙げられる。これより、航空機の音を用いることにより小型マイクなどのデバイスを搭載した野生動物の位置情報の同定を試みる。具体的な概念図を図 3.1 に示す。用いるリソースは上空を飛行する航空音、野生動物に搭載されたデバイスのマイクで録音された環境音、そして固定マイクで録音された環境音の三点である。

前提条件として、5km ほどの間隔で置かれた外部との通信環境が整えられた固定マイクを 3 台設置しているものとする。これらの固定マイクは周囲の環境音を記録する。また、その範囲において先にも挙げたセンサやデバイスが搭載された野生動物が行動しているものとする。そして、そのデバイスに搭載されているマイクにより環境音を録音する。さらに、航空機の位置情報は Flightradar24⁷ と言う web サービスを通して 24 時間インターネット上で公開されている。これを用いることにより、航空機の緯度経度などの情報は既知であるものとする。この条件で環境音を記録すると、航空機が観測地帯において上空を通過し、そのジェット音が記録された際に、音速の到達時刻の観点から録音された音源の記

⁷ Flightradar24.com - Live flight tracker! - <https://www.flightradar24.com/>

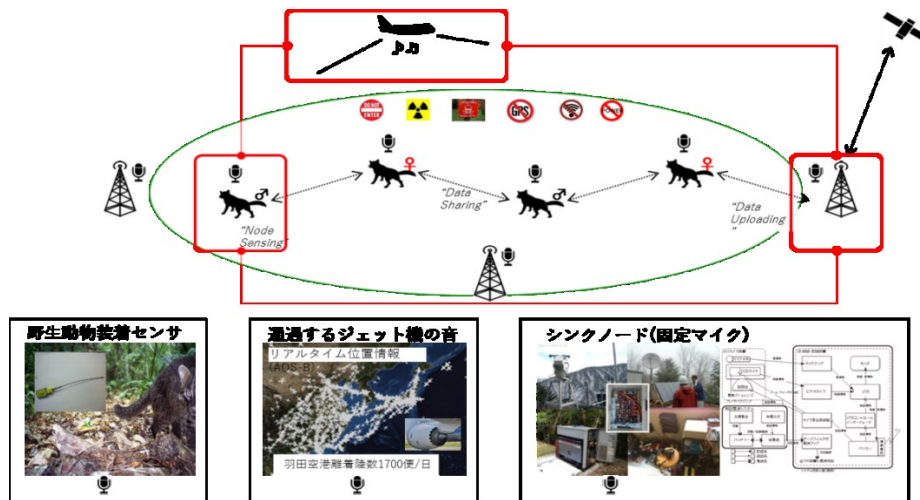


図 3.1 提案システム概念図⁸

録時刻にズレが生じることが考えられる。このズレの開始時刻と三点測位を利用することにより、最終的には野生動物の位置を同定する。以下で具体的な測位手法について説明する。図 3.2 はその手法についての概念図である。まず図 3.2 の(a)では航空機から固定マイクまでの距離の測り方を示している。航空機の位置に関しては既知であり図上では (x, y, z) 、固定マイクの位置を $A(x_A, y_A, z_A)$ とあらわす。図を見ても分かるように、航空機から固定マイクまでの距離は

$$d_A = \sqrt{(x - x_A)^2 + (y - y_A)^2 + (z - z_A)^2}. \quad (3.1)$$

である。次に、航空機から野生動物までの位置を算出する。図 3.2(b)上での t_W は野生動物、 t_A は固定マイクにおけるジェット音の記録開始時間である。また、野生動物の位置を $W(x_W, y_W, z_W)$ とあらわす。そうすると、野生動物までの距離は式変形を行うことで

$$d_W = (340 \times |t_W - t_A| + d_A). \quad (3.2)$$

と表現できる。航空機から野生動物までの距離を得られた後には、固定マイクから野生動物までの距離 q_A を求める。図 3.2(c)がそれをあらわす。航空機の高度を h とする。この際に、固定マイクと野生動物における録音の高度差は考慮しないものとする。航空機直下から固定マイクまでの距離を d_A とすると q_A は

$$q_A = \sqrt{(d_W^2 - h^2)} - d_A. \quad (3.3)$$

⁸ 図は中川ら [5]を基に画像加工を施し作成

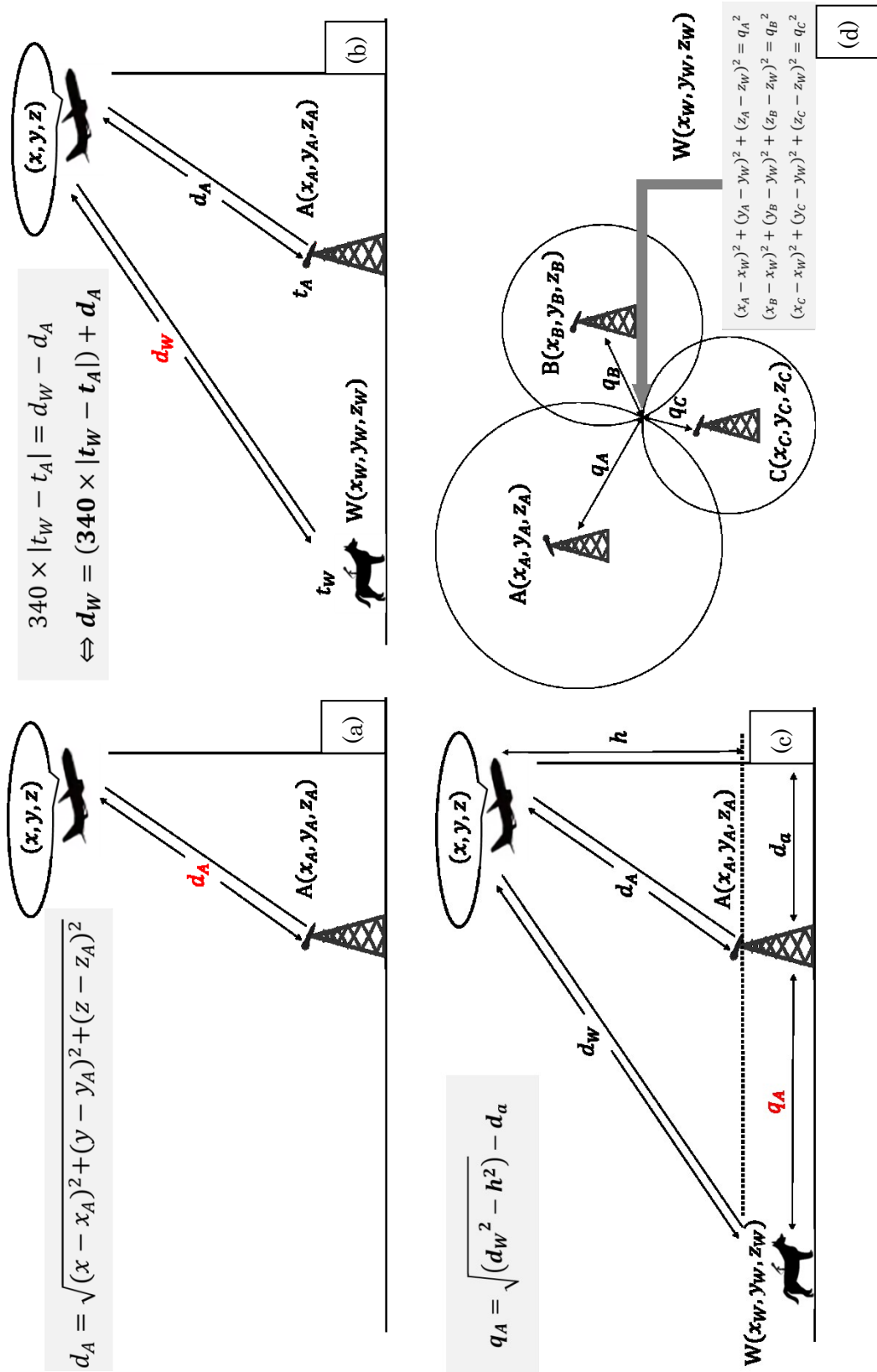


図 3.2 測位手法の概念図

となる。他の固定マイク 2 点でも同様の操作で q_B および q_C を得ることができる。こうして得られた三点を用い三点測位を行う。図 3.2(d)ではその手法についての説明である。こうして、これまでの操作により得られた値を用い、三円の交点を求めることによって未知点である野生動物の位置座標 $W(x_W, y_W, z_W)$ を割り出すことが可能となる。

実際に、調査員が現地で航空音の通過時刻などを記録し、この手法を基に主計算で検証してみたところ、真値との誤差 900m で位置同定を実現している（未発表）。しかし、この推定は調査員による完全手動計測結果によるものであり、1.3 節でも挙げたよう最終的には全自動化を行う必要がある。それを実現するためには、録音された音源内で航空音を自動的に検知できる必要がある。

3.2 音声認識と機械学習を用いた航空音の自動分類への試み

前節において航空音の自動認識を行う必要があることを挙げたが、本節ではそれを実現するために使用した技術や実現へのアプローチについて記す。また、関連研究でも取り上げたよう、今回は音声認識に関しては、環境音の認識について有効な特徴量と考えられる MFCC を用い、分類を行う際に様々なカーネルを用い比較を行える分類器として SVM を用いることにした。

3.2.1 サポートベクトルマシン (SVM)

SVM はクラス分類などを行う上では近年主流となった教師あり学習モデルである。SVM はマージン最大化と言う基準により、与えられた 2 クラスの判別曲線を学習する。

図 3.3 にマージン最大化の概要図を示す。緑色の点と青色の点はそれぞれ分類を期待するクラスの訓練データである。マージンとはこのクラスを分離超平面に写像した際の分類境界と訓練データの最短距離である。分類境界とは図 3.3 において赤色の線であり、分類境界と訓練データの図 3.3 におけるサポートベクトルにより分類境界は一意に定まり、SVM はこのマージンが最大になる分類境界が最も汎用性の高いものとされる [26]。図 3.3 では線形モデルを想定しているが非線形モデルにおいても同様の手法を取る。その際には、計算量が膨大になることをカーネル法と呼ばれる手法を用いることにより避けることが可能である。また、二値分類問題における線形 SVM は(3.4)のように定義される [26]。

$$y(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b. \quad (3.4)$$

ここで、 \mathbf{w}^T と b はパラメータであり、 $\phi(\mathbf{x})$ は特徴空間への写像関数である。なお、 \mathbf{x} は $\mathbf{x}_1, \dots, \mathbf{x}_N$ と N 個の入力ベクトルである。これら、 \mathbf{w}^T および b はマージンを最大化するような値であり、これはラグランジュの未定乗数法を用いることにより定まる。ここで、線

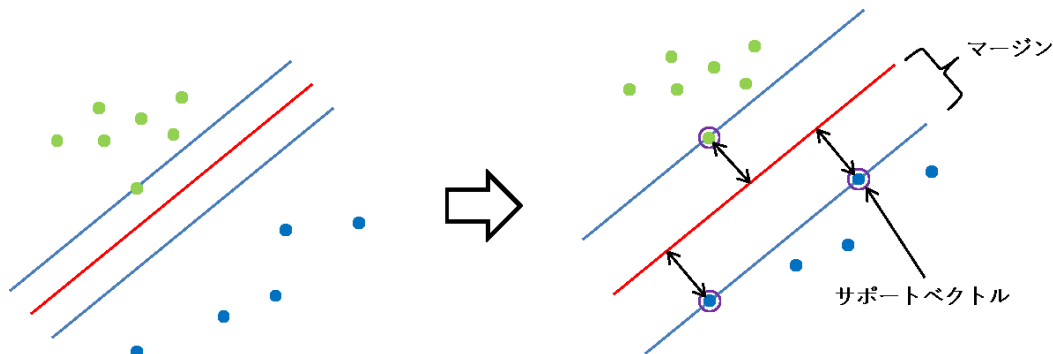


図 3.3 マージン最大化の概念図⁹

形モデルは計算量も少なく高速にモデルを形成できる手法として期待されるが、線形であると分類できるデータに限度があることが想定される。しかし、非線形のモデルで識別を行おうとすると計算量が膨大になる問題が生じる。そこで、カーネル関数を用いた非線形 SVM(3.5)を用いることによりその問題は解決することができる [26]。

$$y(\mathbf{x}) = \sum_{n=1}^N a_n t_n k(\mathbf{x}, \mathbf{x}_n) + b. \quad (3.5)$$

(3.5)において a_n および b は(3.1)と同様にパラメータである。 t_n は(3.1)における N 個の入力ベクトル \mathbf{x} に対応する目標値 t_1, \dots, t_N ($t_n \in \{-1, 1\}$)である。そして、 $k(\mathbf{x}, \mathbf{x}_n)$ をカーネル関数と呼ぶ。通常的手法では計算量がかかる二次計画問題を、双対表現を用い双対問題に置き換えることでモデルをカーネルで表現（カーネル置換と呼ぶ）できるようにし効率化している。言い換えれば、分類境界を導出する際に写像関数同士の内積をとる必要がある。その際に高次元ベクトルであった場合には計算量がかかると予想されるが、この演算を別の関数に置き換えることにより避ける手法がカーネル法である。

カーネル関数には種類がいくつかある。以下に列举する。

カーネル名	関数	
線形カーネル (linear kernel)	$k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}'.$	(3.6)
多項式カーネル (polynomial kernel)	$k(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^T \mathbf{x}' + c)^M. (c > 0)$	(3.7)
ガウスクーネル (RBF kernel)	$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\ \mathbf{x} - \mathbf{x}'\ ^2}{2\sigma^2}\right).$	(3.8)

⁹ 図は Bishop ら [26]を基に作成

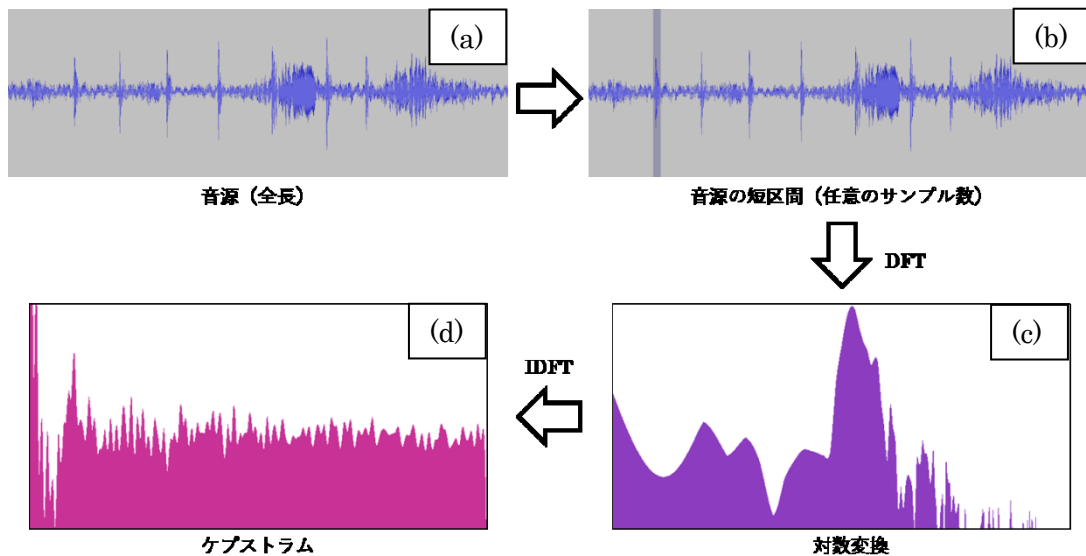


図 3.4 ケプストラム抽出の流れ

各関数において \mathbf{x}, \mathbf{x}' は入力ベクトルであり，多項式カーネルにおける c は定数．また， M は次元数を表す．ガウスクーネル (RBF; radial basis function) における σ はパラメータである．また RBF カーネルを以降は(3.8)式における $-1/2\sigma^2$ を γ で置き換えた $k(\mathbf{x}, \mathbf{x}') = \exp(\gamma\|\mathbf{x} - \mathbf{x}'\|^2)$ と表現する．

3.2.2 メル周波数ケプストラム係数 (MFCC)

MFCC はそもそもケプストラムと言った特徴量に処理を施すことで得られる値であり，ケプストラムは，人の発話内容を認識するために広く用いられている特徴量である．このケプストラムはMFCCと似ているがこの特徴量では人の聴覚特性に対して考慮をしていない点が挙げられる．人の聴覚特性とは，低～中周波の音に対しては分解能が高く，高周波の音に対して低いことである [27]．

図 3.4 では音声信号からケプストラムを取得する過程を示している．特徴量を得るまでの流れは，入力音源において任意のサンプル数を離散フーリエ変換 (DFT; discrete fourier transform) する．その結果に対してさらに対数変換を行う．こうして得られるスペクトラムが図 3.4(c)である．この得られたスペクトラムに対して，今度は逆離散フーリエ変換 (IDFT; inverse DFT) を行うことで図 3.4(d)のようなケプストラムを得ることができる．ケプストラムを任意のサンプル数で得ている理由は，このサンプル数 (フレーム長) をスライドして抽出することにより入力音源におけるケプストラムの体系を得ることができるからである．この出力された体系を特徴量として用いたものが MFCC となる．なお，特徴

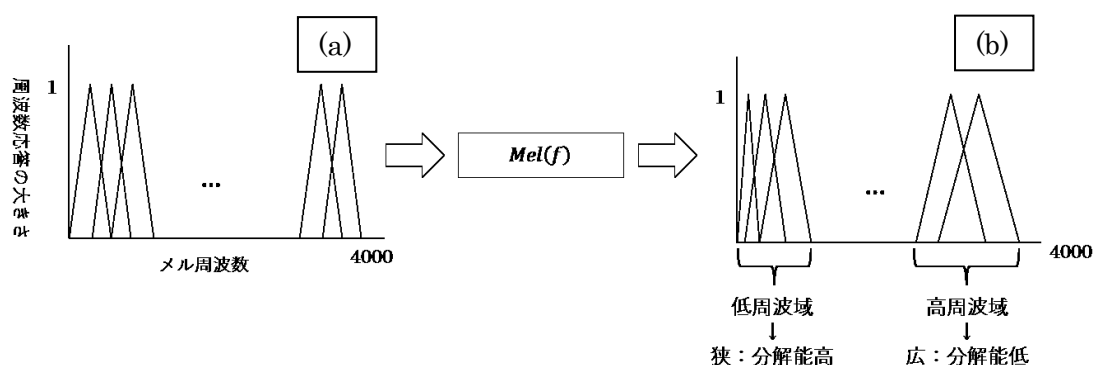


図 3.5 メルフィルタバンク作成の過程¹⁰

量は多次元で与えられ、そのうち最初の 0~12 の計 13 次元を用いることが多い。

上述したように、人の耳は高周波な音ほど捉えにくく、低~中周波な音は捉えやすいと言う特徴があり、この人間の聴覚特性を考慮したような特徴量としてメル尺度というものがある。このメル尺度をケプストラムに対してフィルタを施した特徴量を **MFCC** と呼ぶ。周波数 $f[\text{Hz}]$ におけるメル尺度 $Mel(f)$ は以下の式(3.9)により算出することが可能である [28]。

$$Mel(f) = 1127 \ln \left(1 + \frac{f}{700} \right). \quad (3.9)$$

MFCC の特徴量を得るまでの流れを記す。図 3.4 の手法によってケプストラムを得た後に、メル周波数を求める。メル周波数とはメル尺度により換算された周波数であり、メル周波数軸上に等間隔で配置した図 3.5(a) のような三角窓をフィルタバンクにかけることで求めることができる。フィルタバンクとは、バンドパスフィルタを複数並べたものである。図 3.5(b) を見ても分かるように、低周波数領域では細かく、高周波数領域では荒い、人間の聴覚を考慮した三角窓ができていることが分かる。これは、(3.9) が対数式となっていることからこのような形が得られる。これにより、得られたメルフィルタバンクを入力音源のスペクトルにかけ、最後に離散コサイン変換 (DCT; discrete cosine transform) をかけることにより MFCC の特徴量を得ることができる。

3.3 航空機ジェット音をはじめとした環境音の種々たる特徴

今回、航空音を自動識別するにあたり、我々は航空機のジェット音に特徴的なスペクトラムが存在しないかを検証することにした。検証するにあたり、以下 2 箇所での環境音を用いることにした。

¹⁰ 図は Han ら [28] を基に作成



図 3.6 環境音測定所イメージ¹¹

- オンライン上で取り扱うことが可能である環境音 (Cyberforest¹²)
- 東京大学秩父演習林周辺における環境音

最初に、一つ目の「オンライン上で取り扱うことが可能である環境音」について説明を行う。これは、東京大学サイバーフォレスト研究チームが提供する web サービスであり、森林における環境音を 24 時間誰でも聴くことが可能である。また、日付毎の環境音がアーカイブされている点が特徴である。我々は一つ目の音源として、このサービスを通じ航空機が頻繁に通過すると考えられる千葉県船田池における環境音を実験に用いることにした。

二つ目は、東京大学が所有している秩父演習林での環境音である。厳密には二箇所での環境音であり、栃本と矢竹と呼ぶ箇所において音源の録音を行った。秩父演習林における音源を利用した理由としては、実用を考えた際に想定される環境は人間が存在しないような場所だからである。栃本の周辺は県道と山に挟まれた場所であり、定期的にバイクや自動車と言った乗り物が通過することと野鳥が鳴く環境であるが、矢竹は白泰山の奥地であり、かつ東京大学の演習林と言う背景から一般車輛の乗り入れが禁じられている。これよ

¹¹ 図 4.6(a)に関しては Cyberforest 内より引用

¹² Cyberforest Links - <http://cyberforest.nenv.k.u-tokyo.ac.jp/>

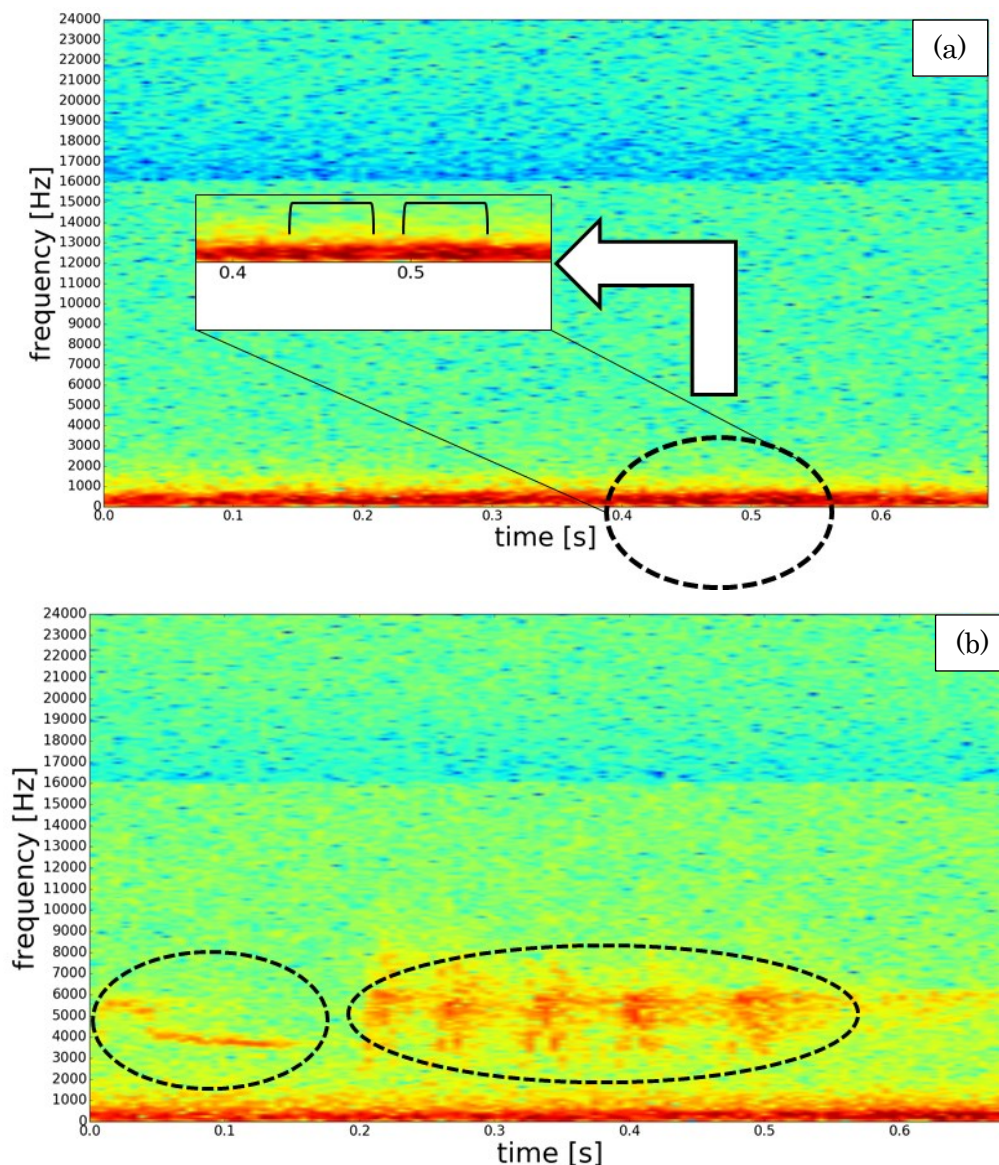


図 3.7 船田における航空音（上）および野鳥音（下）のスペクトラム
（記録形式：mp3）

り、乗り物の音源は記録されないような環境となっている。その点において矢竹の環境は実用を踏まえた環境に近いであろうと言える。また、もう一つの理由として Cyberforest より提供されている音源は非可逆音源であり高周波帯の音源はカットされている観点から、可逆音源であり高周波帯も情報を残す音源を収録する目的がある。高周波帯における情報を残す理由については、後述する。この点については、矢竹では長時間の収録が行程の都合上行えなかったため、長時間の音源記録が行えた栃本において補う。

各測定場所については図 3.6 を参考にされたい。便宜的に以降では 3.6(a)を船田、3.6(b)

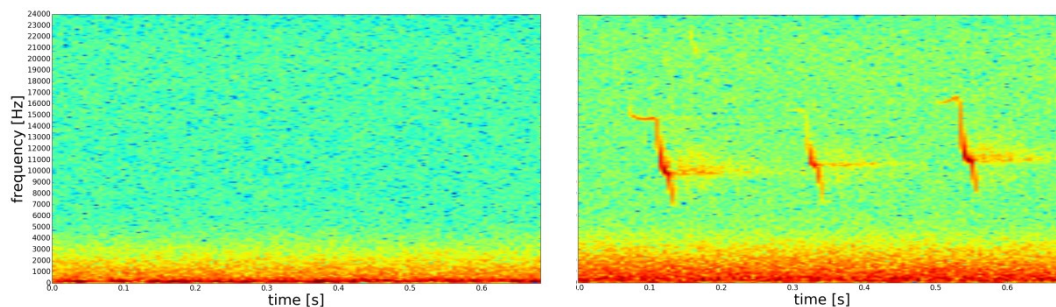


図 3.8 桧本における航空音（左）および野鳥音（右）のスペクトラム
（記録形式：wav）

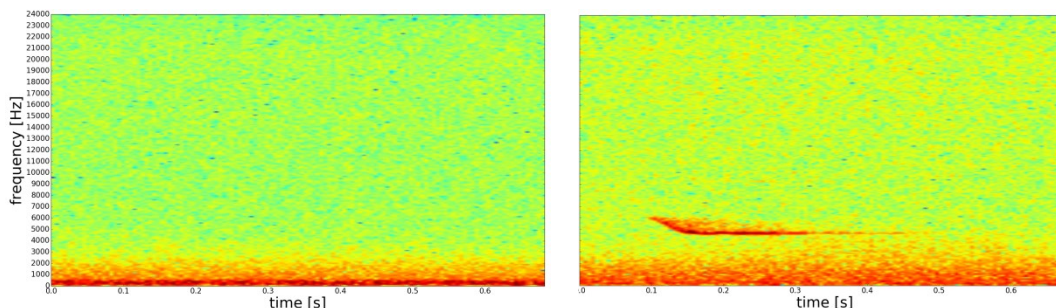


図 3.9 矢竹における航空音（左）および野鳥音（右）のスペクトラム
（記録形式：wav）

を桧本、3.6(c)を矢竹と呼ぶ。

はじめに、船田の音源を解析してみたところ、航空機音と野鳥音のそれぞれにおいて特徴的なスペクトラムが見受けられた。図 3.7 はその一部を示す。ここで、図 3.7(a)は航空音のスペクトラムであり、図 3.7(b)は野鳥音のスペクトラムである。また、(a)と(b)における色は音圧の強さであり、赤に近い色ほど強く、青に近いほど弱い入力であることを示す。

この図 3.7(a)では少し分かりにくいですが、航空機音は 500Hz 付近において通過時に強い音圧（黒括弧で示している箇所）が得られることがわかる。また、野鳥音に関しては、2kHz 以降の地帯において特徴的なスペクトラムが存在することが判明した。

これらの結果から、航空音は 500Hz 帯を、野鳥音は 2kHz 帯以降に特徴的なスペクトラムが見受けられたと判断したいが、研究背景でも述べたように、航空音は高周波帯において特徴的なスペクトラムが出る傾向があると示されていることや、経験則的に航空音は通過時に高音が響く印象がある。これらの事例から、今回解析した結果では偶然にも高周波帯において特徴的なスペクトラムが存在しなかった可能性が否めない。

そこで、船田の音源の問題点を踏まえ我々は可逆記録方式で収録した桧本と矢竹の音源のスペクトラムを解析することにした。図 3.8 および図 3.9 は、桧本と矢竹における航空音

表 3.1 各音源の記録形式など詳細

	船田	栃本・矢竹
記録形式	MPEG Layer-3 (mp3)	RIFF (wav)
サンプリングレート	48kHz	48kHz
ビットレート/深度	VBR (可変)	16bit
チャンネル数	2ch	2ch

と野鳥音におけるスペクトラムの一部である。録音した音源を解析してみたところ、航空音において高周波帯における特徴的なスペクトラムはやはり見受けることができなかった。しかし、野鳥音においては船田と共通して双方とも 2kHz 帯以降において特徴的なスペクトラムを確認することができた。このスペクトラムは野鳥の種類によって異なることが図から判断できる。この結果より、以降は航空音については高周波帯において特徴的なスペクトラムが無いと仮定し、航空音自動認識のシステムを構築してゆく。また、収録した音源から野鳥音は 5kHz 帯付近において多く見受けられたことより、以降のシステムでは 0～5kHz の音を活用するものとする。

各地の音源における記録形式などの詳細を表 3.1 に示しておく。次章にて行う実験において音源はこの形式で統一している。また、船田での音源はサンプリングレートが 48kHz ではあるものの、上述した通り高周波域は存在していない。

第4章

提案システムの評価実験

4.1 実験の概要 —航空音識別手法—

今回、収録した音源を分類するにあたり前章でも取り上げた航空音と野鳥音に加え、バイク音、環境音（何も音が発生していない状態）の計4つの音源をクラスとして用いることにした。それぞれの音源は手動で分割し、1クラスあたり100個と計400個のデータを用意した。その際に、音源が確かにその音なのかどうかと言う真値に関しては人の耳で判別し、ラベル付をした。なお、一つ一つの音源におけるサンプル数（フレーム数）は高速フーリエ変換を用いる理由から2の冪乗数である必要がある。それより今回はサンプル数を $2^{15} = 32768$ と約0.683秒分とした。

これら1クラスあたり100個のデータを、教師データ75、予測データ25とランダムに分割しSVMでは3種類のカーネル（Linear, Polynomial, RBF）を用い分類を行った。なお、以降では分類の際に教師データに音声処理を施さず分類器にかけることを単純処理、認識率向上を図りMFCCにかけ分類を行う手法をMFCC処理と呼ぶ。単純処理における特徴量は、音源をサンプル数512で高速フーリエ変換にかけた後に、全フレームにおける音圧の平均値を特徴量とした。MFCCに関しては、各フレームで得られた13次元の特徴量を、単純処理と同様に全フレームの平均を特徴量とした。

図4.1は単純処理の具体的なイメージ図をあらわす。赤線は航空音、灰線は野鳥音の特徴量を表現している。図3.7でも挙げたように、航空音は500Hz帯でのゲインが強く、野鳥音では2kHz帯以降で強く見られる。この図4.1からは航空音と野鳥音における特徴量（赤と灰の直線）の位置に差は無いように見える。しかし、この場合であると航空音の特徴量は0.0229で野鳥音は0.0148と航空音の数値が大きく出ている。これは音圧が特徴量に影響するため航空音の値が高く出るとは妥当な結果であると言えよう。

この後に、分類器に特徴量を渡すことにより学習を行うのであるが、直接、分類器に値を渡すとLinearやPolynomialカーネルを用いている際には行列の内積計算や高次元演算を行う必要性が生じ、学習速度に多大なる時間が生じる問題がある。そこで、数値演算を

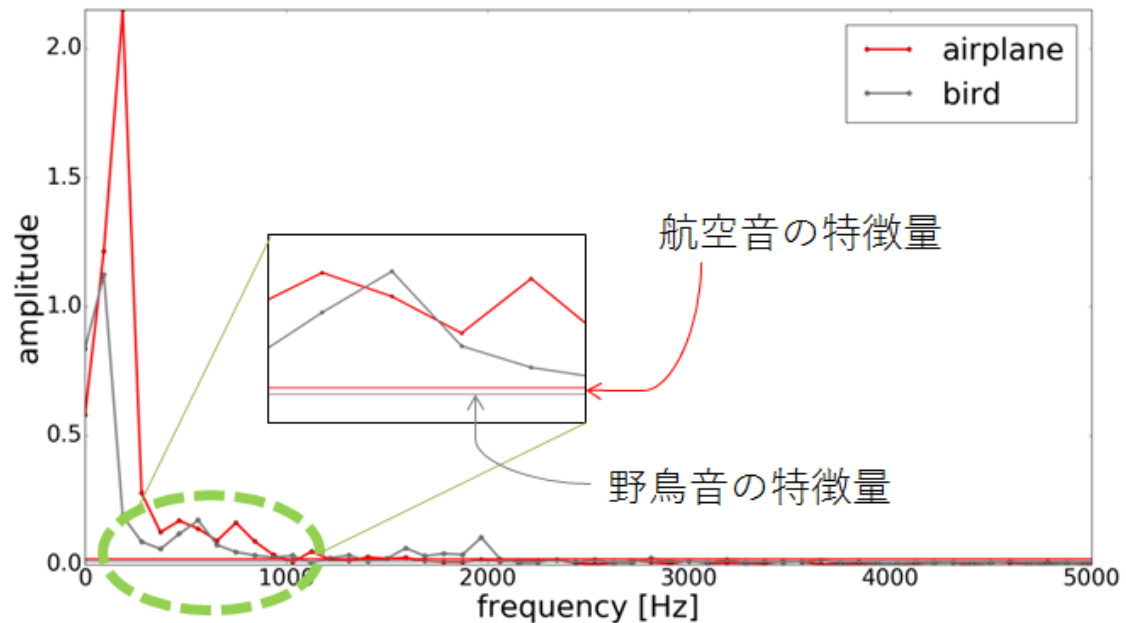


図 4.1 単純処理における特徴量

高速にするため，学習機に特徴量を渡す前に値の標準化を行っている．具体的には，上記の手法で得られた特徴量を平均 0 分散 1 となるようにすることで標準化を施した．また，分類結果を示すには混同行列で可視化し，分類器自体の性能に関しては ROC 曲線を用いることにより評価を行った．

4.2 実験システム・条件

実験をする際においてプログラム言語は機械学習のライブラリが豊富である Python (バージョン 2.7.10) を用いた．また，音声処理および機械学習を行う際には他にもいくつかのライブラリを用いている．それぞれのパッケージ名とバージョン情報について挙げる．MFCC の特徴量を得るまでの演算は scikits の talkbox (バージョン 0.25)，FFT は scipy (バージョン 0.17.1) を用いることにより行う．SVM は sklearn (バージョン 0.17.1) を用いることにより分類や，予測などの処理を行う．また，行列演算の処理を簡略化するために numpy (バージョン 1.10.4) を用いている．

図 4.2 はそのプログラムのフローチャート図である．以下ではプログラムの流れについて順に説明してゆく．プログラムが開始すると，任意のディレクトリに存在する音声すべてを対象に読み込む．その後，過去に入力音声のキャッシュファイルが作成されているかを確認する．キャッシュファイルを作成する理由は，プログラムの実行毎に FFT や MFCC と言った音声処理を行うのは処理に時間がかかることが想定されるからである．また，そ

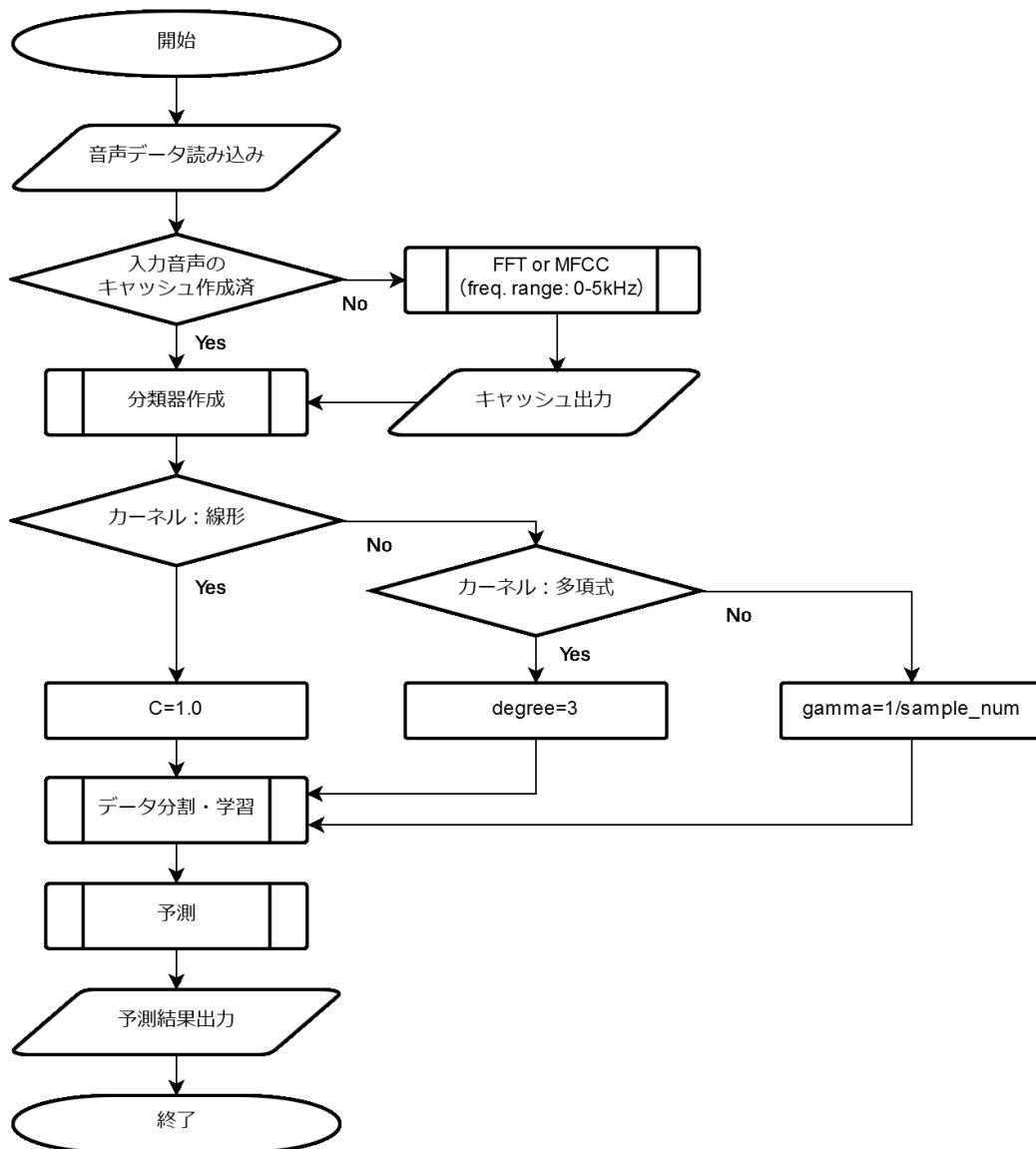


図 4.2 航空音自動認識システムのフローチャート図

の処理を行った結果は毎回同じ結果になることから、音声処理を行った際にその結果を書き出しておくことで次回からはその音源を教師データや予測データとして扱うことが容易となるメリットが挙げられる。このような理由からキャッシュファイルが作成済みかどうかであることを確認する。キャッシュファイルが書き出しされていなかった場合について進む。この場合は、FFT か MFCC の音声処理が入力音源に対して行われる。行われる音声処理に関しては FFT か MFCC のいずれかを選択したファイルにより決まる。その際、上述したように 0~5kHz の周波数帯のみを扱い、それ以上は切り捨てている。処理が終了した後は、結果をバイナリデータとして任意のディレクトリへと出力する。以上が音声処理からキャッシュファイル出力までの工程である。

次に、SVM を行う際に用いる分類器を設定する。なお、入力音声のキャッシュファイルを過去に作成していた場合は直接この処理へと遷移する。分類器は、3.2.1 節で取り上げたように、Linear (線形) と Polynomial (多項式)、そして RBF (ガウス) カーネルの三種類を扱い比較および評価を行う。この際に、どのカーネルを用いるのかと言う判定を行う。判定は、コード内で事前に記述しておくことにより行う。また、カーネルを使用する際にはパラメータを選択する必要がある。なお、本研究内ではパラメータの調整を必要とすることなくどれほどの性能を得ることができるのかを評価したい理由もあり、Python 上で用いられるデフォルトのパラメータを用いている。図上では便宜的にカーネル毎におけるパラメータの値を示している。すべてのカーネルには C というパラメータが設定される。 C とは誤り許容率であり、この値が大きければ大きいほど誤分類を許さないパラメータとなる。図 4.2 にもあるよう、今回この C の値が 1.0 となる。次に多項式カーネルの場合は、degree というパラメータを設定する。この degree は多項式カーネルにおける次数であり、(4.4)式における M に対応する。つまり、値が大きければ大きいほど非線形な分類に対応したカーネルとなることが考えられる。最後に、RBF カーネルの場合は、gamma というカーネルが設定される。(4.5)式における γ がそれに対応する。この gamma はデフォルトでは 1 クラスにおけるサンプル数の逆数がパラメータとして用いられる。今回のケースであると、例えば、航空音のサンプル数は 100 なので gamma のパラメータ値は 0.01 となる。なお、 C に関しては多項式および RBF カーネルにおいても設定されているが、図上では割愛している。

これらの操作が終了した後に、データを教師データと予測データに分割を行う。本研究では全体の 25% を予測データ、残りの 75% を教師データとした。なお、本来は分割処理のたびに教師データと予測データがランダムに入れ替わるようになっているが、評価を行うにあたり分割の規則は一定になるようにした。言い換えれば、毎回実行するたびに選択される教師データと予測データは必ず同じになるということである。この分割が終了した後に分類器において学習を行い、分類予測を行う。以上が航空機音分類システムの流れである。

次節では実験の結果および評価について記す。入力音源は先述したよう船田、秩父演習林周辺における栃本、矢竹の計 3 箇所を使用し、分類を行った。また、上述したが森林内において考えられる音は今回分類を行う航空音をはじめ、野鳥の音、環境音 (航空機や野鳥音などが入っていない音源)、そして乗り物の音などが挙げられる。実験では、ここで挙げた計 4 種類の音源を用い分類を行う。なお、先にも挙げた通り矢竹で記録した音源は乗り物が入れないような環境であったため、航空音、野鳥音、環境音の 3 クラスでの分類結果となっている。また天候の条件は晴天とし、かつ、音源内では風や人の声に起因するノイズがないものとする。ただし、航空機音と野鳥音が混在したような音源は入力音源として認めている。

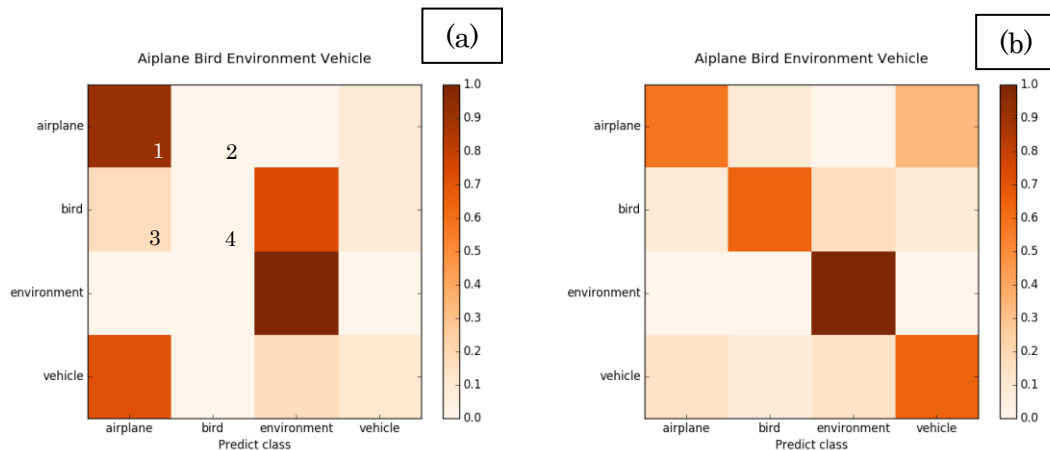


図 4.3 分類結果についての精度を表す混同行列

4.3 実験結果

結果を示す前に、まず今回の実験結果を表す混同行列の見方についての説明を行う。図 4.3(a)および(b)のような行列を混同行列と言い、縦軸が正解クラス、横軸が予測クラスである。例えば、野鳥音の分類に関して判断をしたい場合は、縦軸の野鳥音の項目に対し横軸の野鳥の項目を対応させれば良いことになる。また、色の濃度は分類結果の正解率であり、正解率は以下のように定義される [29]。

$$accuracy(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} 1(\hat{y}_i = y_i) \quad (4.1)$$

(4.1)における y は予測結果のラベルであり \hat{y} は正解ラベルである。また、 $n_{samples}$ は予測データの数である。つまり、予測結果と正解ラベルの値が等しければ 1 の和を取り、最終的に予測データ数分で割った値である。これは、予測データの内いくつかのデータが分類できたのかと言う割合を示している。その際に 0% (分類精度: 低) に近くなればなるほど白い色、100% (分類精度: 高) に近くなればなるほど茶色い色となる。このことから図 4.3(b)のように対角線上に色が濃い図が得られることが理想的である。対して、図 4.3(a)のようなグラフは分類性能が低いことになる。また、項目については上から順に航空機、野鳥音、環境音、バイク音である。なお、SVM は通常 1 対 1 のクラス分類を行うことを想定した分類器でありため、本研究では多クラス分類において用いられる 1 対他分類手法で多クラスの分類を行う。その際には、ランダムに分類した場合だと分類できる確率は 50%となる。また、この実験で用いた各音源のサンプル数は 100 より、二項分布を考えると $B(100, 0.5)$ という分布を取る。それでは、表 4.1 では単純処理、表 4.2 では MFCC 処理における混同行列を示す。

表 4.1 単純処理におけるカーネル毎の分類結果

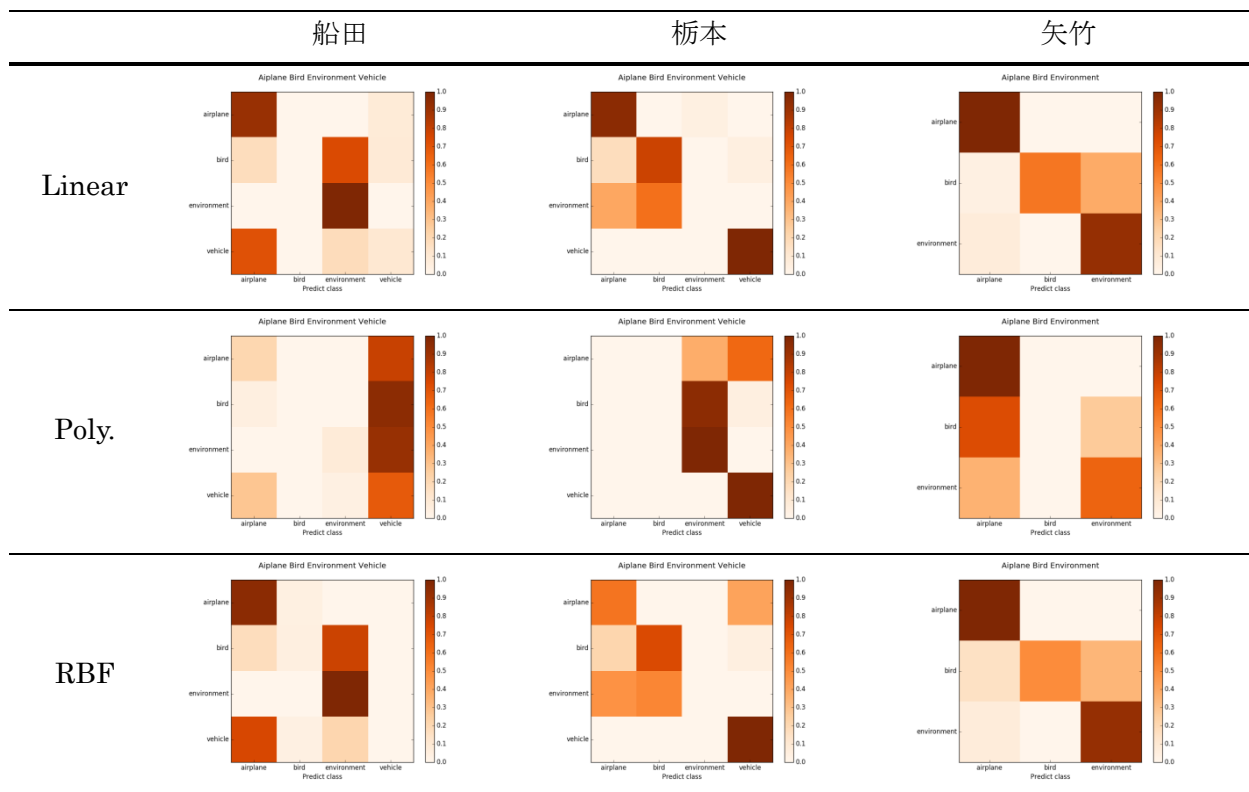


表 4.2 MFCC 処理におけるカーネル毎の分類結果

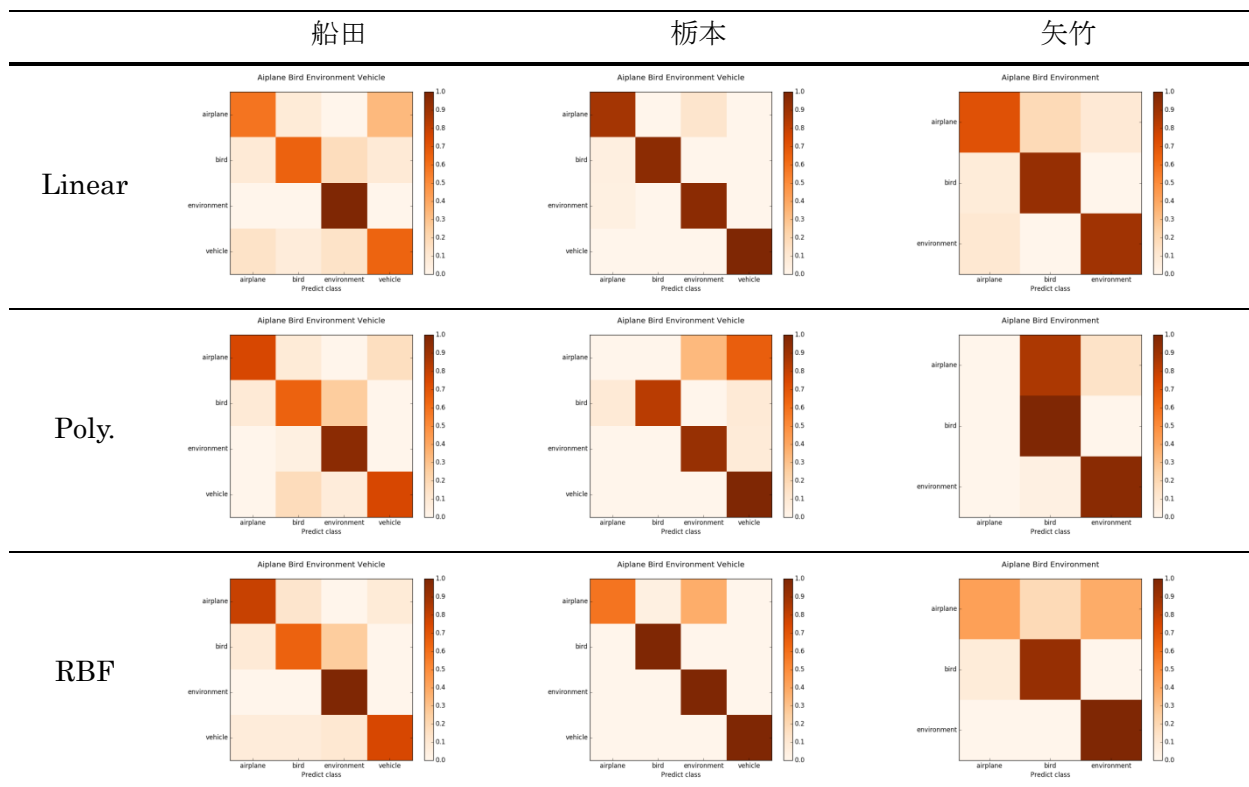


表 4.3 単純処理における分類正解率

	船田	栃本	矢竹
Linear	50.0%	69.0%	82.7%
Poly.	26.0%	53.0%	52.0%
RBF	49.0%	59.0%	80.0%

表 4.4 MFCC 処理における分類正解率

	船田	栃本	矢竹
Linear	72.0%	95.0%	85.3%
Poly.	78.0%	70.0%	70.7%
RBF	80.0%	90.0%	81.3%

表 4.5 ROC 曲線の成り立ちをイメージするためのサンプルデータ

予測確率	実データとの正誤
0.95	True
0.70	True
0.60	False
0.50	True
0.10	False

表 4.3 では単純処理, 表 4.4 では MFCC 処理における分類の正解率を示す. この結果から, 今回の 3 または 4 つの音源下では分類性能が最大で 95.0% という結果を得られた. 特に, 単純処理と比較して MFCC 処理の方が分類精度は向上していることが伺える. また, 使用するカーネルによって精度が異なり, Linear カーネルか RBF カーネルのいずれかの性能が高いように見受けられる. どちらのカーネルを用いるべきかについては, 次節で考察してゆく.

4.4 実験評価

表 4.3 および 4.4 から, 単純・MFCC 処理の双方について多項式カーネルを用いた分類は低い正解率を示すことが判明した. しかし, 線形および RBF カーネルに関してはこれだけでは有効なカーネルかどうか判断しがたい. そこで, 分類器自体の評価を行うことにする. 我々は, 分類器の評価を行うことが可能な ROC 曲線を用いることで分類器自体の評価を行うことにした. ROC 曲線とは, 受信者動作特性曲線 (ROC; receiver operating characteristic) とも呼ばれる. 以下では簡単に説明を行う.

表 4.6 単純処理におけるカーネル毎の ROC 曲線

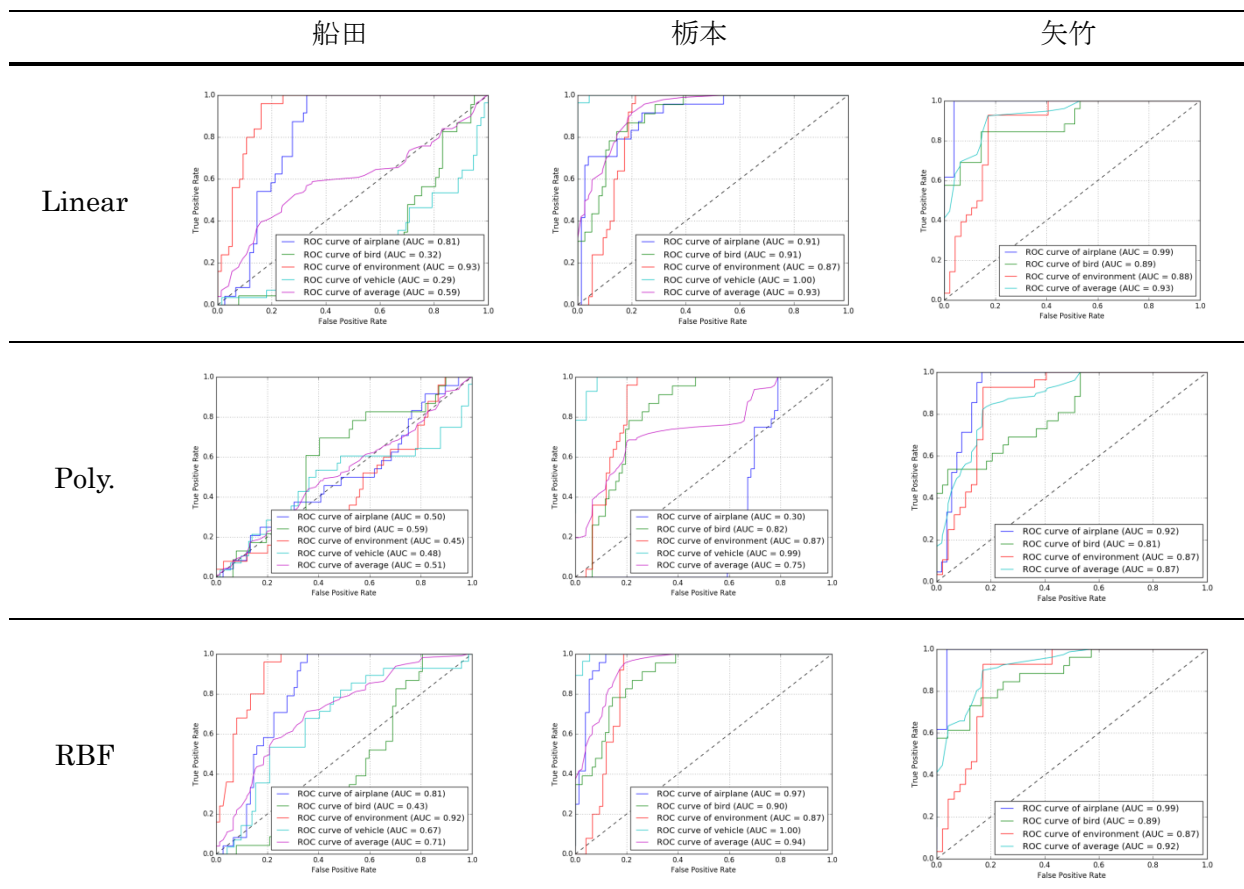


表 4.7 単純処理での ROC 曲線における AUC 値 (航空機と全体平均のみ抜粋)

		船田	栃本	矢竹
Linear	航空機	0.81	0.91	0.99
	全体平均	0.59	0.93	0.93
Poly.	航空機	0.50	0.30	0.92
	全体平均	0.51	0.75	0.87
RBF	航空機	0.81	0.97	0.99
	全体平均	0.71	0.94	0.92

ROC の曲線を描く際には、混同行列における以下の要素を理解しておく必要がある。図 4.3(a)において航空機クラスと野鳥クラスの 2 クラスに着目したとき、一般的に正解クラスが予測クラスと一致した場合を True Positive(TP)と表現する。これは図 4.3(a)-1 がそれに対応する。また、正解クラス以外のクラスが予測クラスと一致した場合は True Negative(TN)と言い、図 4.3(a)-4 が対応する。正解クラスにおいて、予測クラスと不一致だった場合は False Negative(FN)と言い図 4.3(a)-2 が対応する。最後に、正解クラス以外

表 4.8 MFCC 処理におけるカーネル毎の ROC 曲線

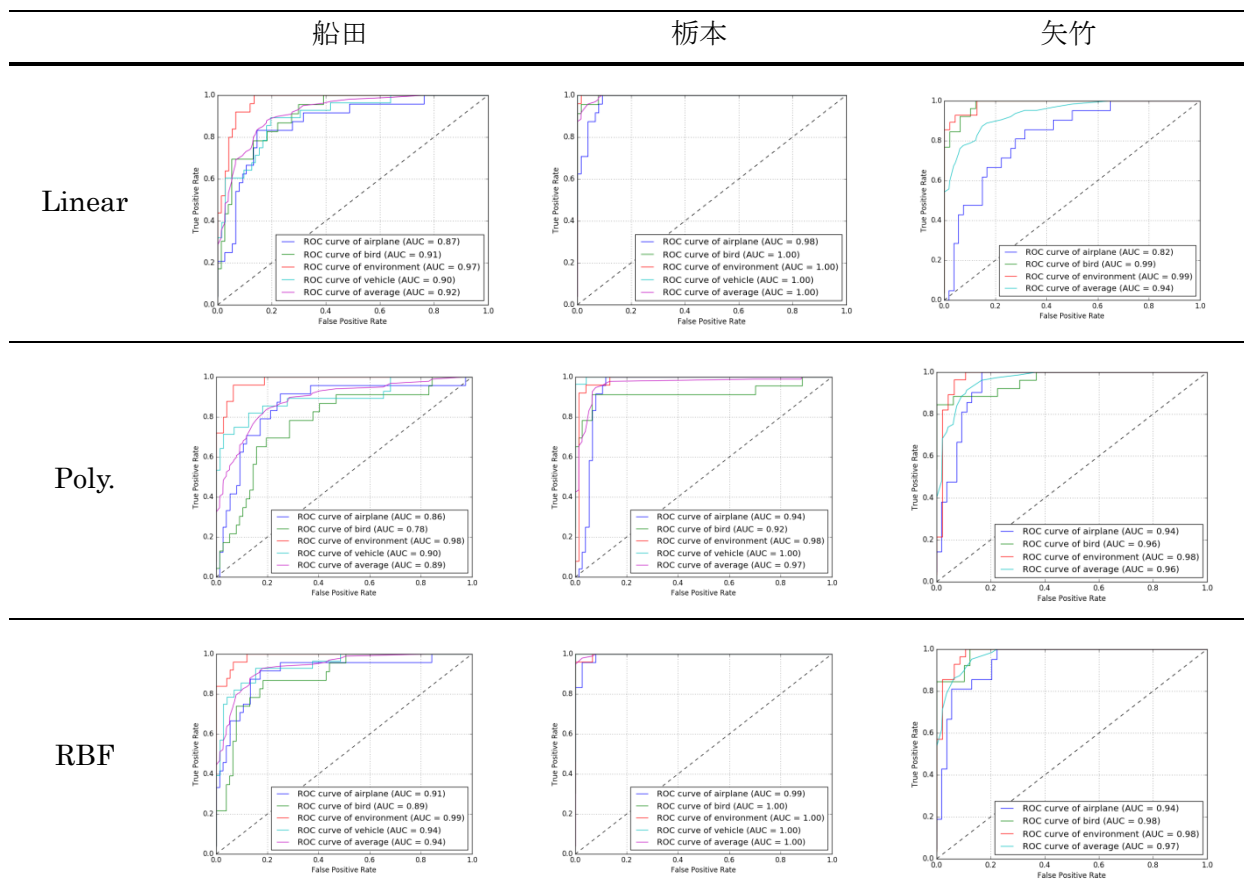


表 4.9 MFCC 処理での ROC 曲線における AUC 値 (航空機と全体平均のみ抜粋)

		船田	栃本	矢竹
Linear	航空機	0.87	0.98	0.82
	全体平均	0.92	1.00	0.94
Poly.	航空機	0.86	0.94	0.94
	全体平均	0.89	0.97	0.96
RBF	航空機	0.91	0.99	0.94
	全体平均	0.94	1.00	0.97

において、予測クラスと不一致だった場合は False Positive (FP) と呼び、図 4.3(a)-3 が対応する。

今回の航空音認識を例に挙げると、音源の分類を行う際に任意の音源がその音らしいであろう確率を、交差検証法を用いることにより算出している [30]。その確率と実際の正誤が例えば表 4.5 のようになったとする。ここで、True は全体で 3 つあり、False は 2 つである。この後、任意の確率においてそれ以上の値は Positive、それ以下は Negative と閾値

を変えながら TP と FP の値を算出する。表 4.5 の場合だと以下ようになる。

- 0.95 以上が Positive ならば TP : 1/3 FP : 0/2
- 0.70 以上が Positive ならば TP : 2/3 FP : 0/2
- 0.60 以上が Positive ならば TP : 2/3 FP : 1/2 …

このように、順に計算を行ってゆく。その計算によって得られた値を縦軸は TP，横軸は FP の値をプロットする。そうして得られたグラフが ROC 曲線となる。また，ROC 曲線の曲線下部の面積を ROC 曲線下面積 (AUC; area under the curve) と呼び，分類器の性能を評価するための指標となる。なお，AUC は 0.5 から 1.0 の値をとり，1.0 に近ければ近いほど，性能の良い分類器であることを示す。

表 4.6 から表 4.9 において，この ROC 曲線を使用した分類器の評価結果を示す。表 4.6 と表 4.8 における青色の線は航空機音，緑色の線は野鳥音，赤色の線は環境音，水色の線はバイク音の ROC 曲線値である。また，紫色の線は全クラスの ROC 曲線値における平均値を表した曲線である。表 4.7 および表 4.9 においては ROC 曲線の航空機の AUC 値と AUC 全体の平均値を抜粋した。この結果から見て分かるように，単純処理では，船田での音源に対して全体平均の値が最大でも 0.71 の AUC 値であったものが，MFCC 処理を通すことにより，同地帯において最大 0.94 と AUC の値が改善された。また，前節で取り上げた Linear カーネルと RBF カーネルの優劣に関しては，この結果に鑑みると RBF カーネルのほうが適していると言えよう。しかし，栃本と矢竹の二地点においては，カーネル次第では単純処理と MFCC 処理の間で遜色ない結果が出ている点に疑問が生じた。この点に関して次節では取り上げる。

4.5 本節における実験での問題点

矢竹に関しては，先述したように限られた時間内でしか音源を録音できなかったことや，栃本に関しても，録音された音源上では数種の野鳥がいたのみであり，航空音や野鳥音のデータに多様性がない音源であることが挙げられる。また，栃本と矢竹における航空音の入力が弱かった背景があることから，バイク音，野鳥音，航空音，環境音の順に音圧が分類において鍵となる単純処理においてもカーネル次第では高い精度を挙げられたものだと考えられる。この他にも，サンプル数が高々 100 と少なく本当に分類器の精度が良いと判断するのは妥当なのかと言う疑問が生じた。これらの疑問を払拭すべく，我々はこれを改善案とし，試してみることにした。具体的には，三地点での音源を混合することでサンプル数を増やし，かつ，航空音のさらなる分類精度向上を図った。特に，航空音には 500Hz 帯に特徴的なスペクトラムがあることから，ローパスフィルタをかけることでさらなる精度向上が図れるのではないかと考えた。これらに関する結果を次章で述べる。

第 5 章

ローパスフィルタによる 提案システムの改善

5.1 提案システム改善案の概要 —フィルタ通した識別—

4.5 節で挙げたように、野鳥音を分離するために我々はローパスフィルタを用い分類精度が向上するのではないのかと言う仮説をたてた。これを確かめるべく、以下のような条件でローパスフィルタを設計した。

- 通過域：500 [Hz]
- 遷移域：1000 [Hz]
- フィルタ回路：FIR フィルタ
- サンプル数：512
- 窓関数：Hanning 窓

この際に設計したローパスフィルタは航空機の音源データのみにかけることでノイズである野鳥音との分離を行う。実際に、その航空音から今回航空音の特徴があらわれると仮定した 500Hz 帯以上にローパスフィルタにかけることでそれ以降の帯域を除去した状態を図 5.1 にあらわす。なお、図 5.1 上の LPH はローパスフィルタを意味する。

また、4.5 節内での実験ではサンプル数の少なさが懸念事項であった。そこで、追加実験内では 4 章の実験と同条件で作成した船田、栃本の音源データを混在させた計 200 個の音源で航空音、野鳥音、環境音、バイク音の 4 クラスを教師データと予測データと分割することによりサンプル数が多くなった際にどのような挙動を示すのかを検証した。また、同一箇所矢竹を加えた計 300 の音源データを航空音、野鳥音、環境音の 3 クラスでの分類も同様に行った。

次節では、その実験結果を示す。

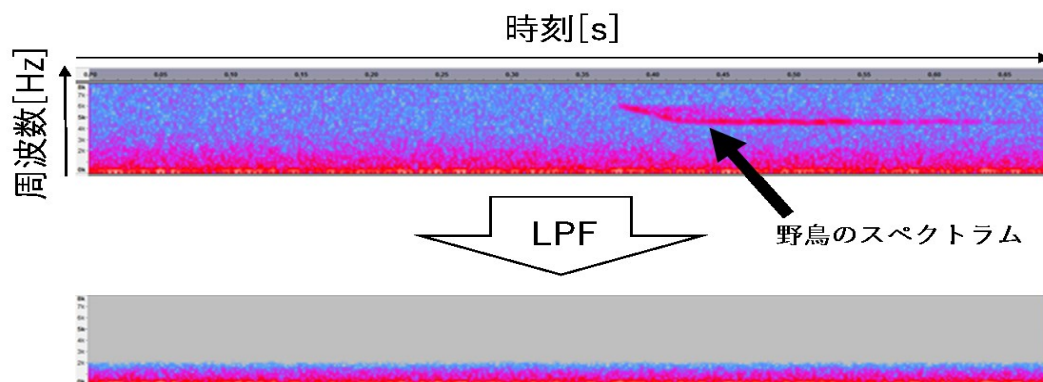


図 5.1 ローパスフィルタによる野鳥音の除去

表 5.1 単純および MFCC 処理における 3 および 4 クラス毎の分類結果

	3-Classes (単純)	3-Classes (MFCC)	4-Classes (単純)	4-Classes (MFCC)
Linear				
Poly.				
RBF				

表 5.2 単純および MFCC 処理における分類正解率

	3-Classes (単純)	3-Classes (MFCC)	4-Classes (単純)	4-Classes (MFCC)
Linear	61.3%	90.2%	49.5%	86.5%
Poly.	35.6%	91.1%	23.0%	84.5%
RBF	56.9%	84.0%	48.5%	86.0%

5.2 改善案を用いた実験結果

表 5.1 から 5.3 にて、ローパスフィルタを通した単純および MFCC 処理の結果を示す。4-Classes というカラム名は船田、栃本の二箇所音源を計 200 個の音源データを用い分類した結果である。3-Class は音源として矢竹を増やし、その分バイク音を減らしたものである。この際、教師データと予測データの分割は 4 章の実験と同じように 4-Classes の教師データは全体の 75% (225 個)、予測データは 25% (75 個) と分割している。3-Classes においても同様の割合で分割を行っている。

表 5.1 および表 5.2 の分類結果を見てわかるように、やはり MFCC 処理をかけた分類のほうが高い精度を得ることができた。三観測所の音源データを混在させることにより、栃本や矢竹のようなスペクトラムの分布が似たような音源ばかりではなくなったことから、綺麗に分類できているような図は得られなかった。しかし、これは今後実際に分類器を用いる際には、今回用意した音源とはさらに異なった音源が入力音源として想定されることから好ましい結果であると言えよう。

各処理についてであるが、単純処理においては 3 と 4 クラスの双方において航空機の分類はローパスフィルタをかけたことから分類精度は高く出ている。しかし、他のクラスにおいて分類精度が芳しくないことが言えよう。表 5.3 を見てもわかるように正解率が最大でも 61.3% とランダムに分類した結果よりも多少良い程度の分類精度となっている。これでは分類機として機能していないと言える。対して、MFCC 処理は野鳥音を除いた計 3 クラスにおける分類は非常に高い精度が出た。野鳥音の分類については他のクラスと比較すると良いとは言えない結果であるが、航空機と同じく野鳥音特化した特定の周波数帯を通すようなフィルタを設計することにより対処できるのではないのかと考えられる。

今回、ローパスフィルタをかけることにより分類精度は向上したように思われるが、実際に分類機の精度を定量的に比較・把握する必要もある。そこで、前章で扱った ROC 曲線を再度用いることによりここでも分類器の評価を行う。

5.3 改善案の評価

表 5.4 および表 5.5 では 3 と 4 クラス毎の ROC 曲線の結果を示す。この表 5.4 の 3 クラスにおける ROC 曲線からは航空機の分類精度は単純と MFCC の双方で良好であるが、環境音と野鳥音の分類は芳しくないことが見受けられる。また、表 5.5 の 4 クラスにおける ROC 曲線からは、単純処理において線形カーネルと RBF カーネルは航空機の分類がよく行えているものの多項式カーネルにおいては AUC の値が 0.39 とランダムな分類以下の結果となっている。これら 4 クラスにおける単純処理は航空機の分類以外に関しては同じような ROC 曲線を象った。

表 5.3 3 クラスにおける双方処理の ROC 曲線

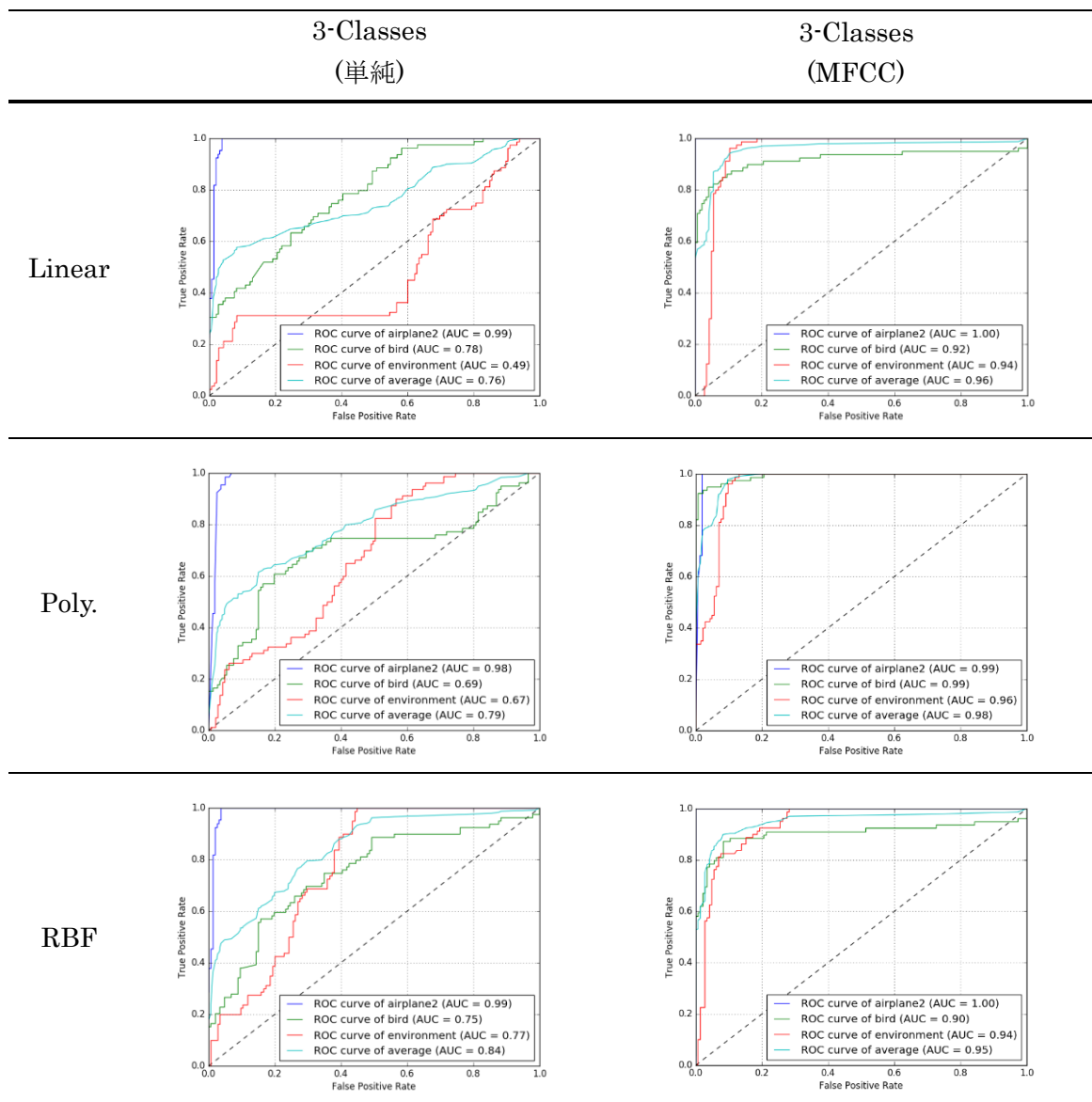


表 5.4 3 クラスにおける双方処理での AUC 値

		航空機	野鳥音	環境音	平均
Linear	単純	0.99	0.78	0.49	0.76
	MFCC	1.00	0.92	0.94	0.96
Poly.	単純	0.98	0.69	0.67	0.79
	MFCC	0.99	0.99	0.96	0.98
RBF	単純	0.99	0.75	0.77	0.84
	MFCC	1.00	0.90	0.94	0.95

表 5.5 4クラスにおける双方処理の ROC 曲線

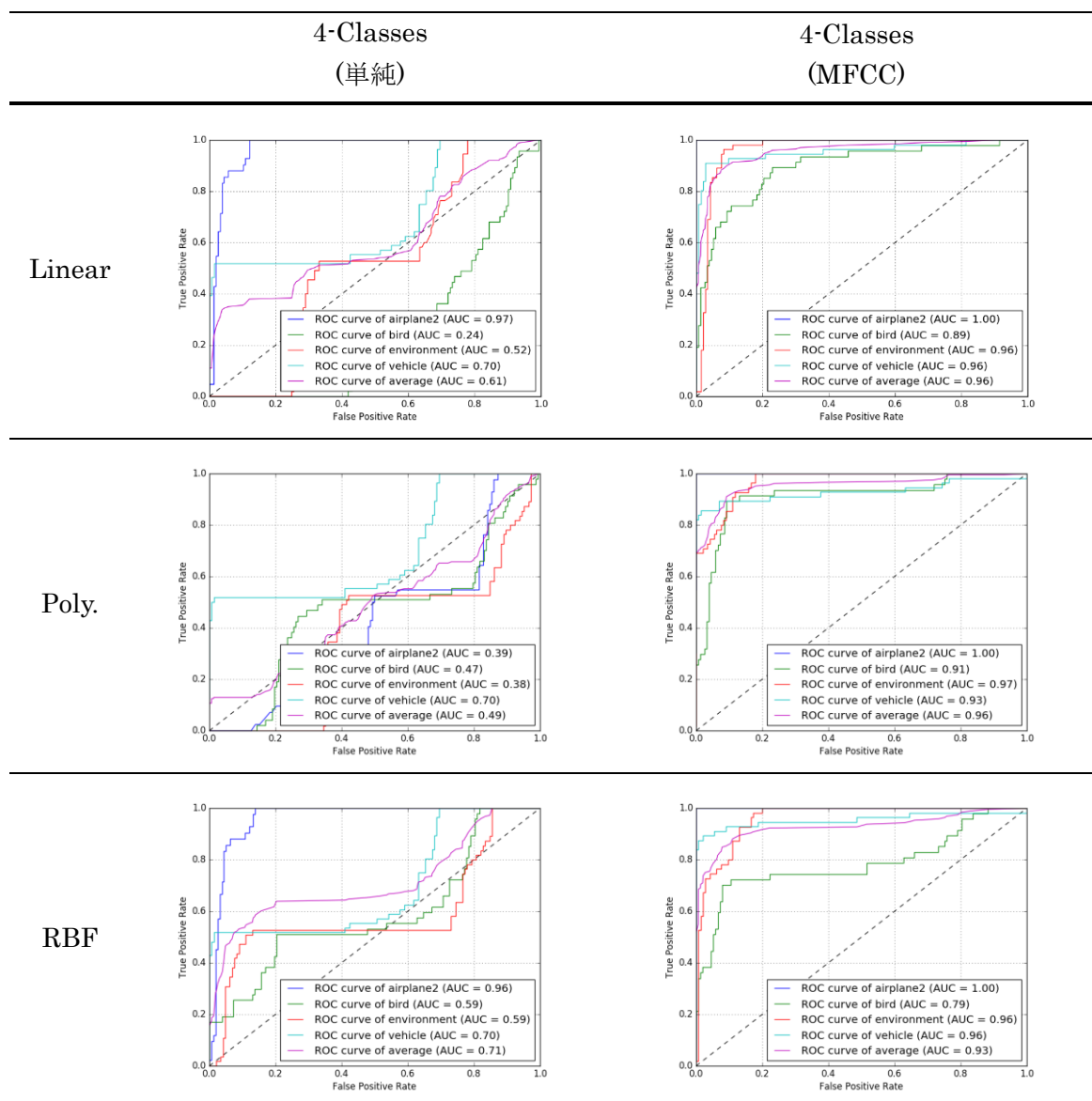


表 5.6 4クラスにおける双方処理での AUC 値

		航空機	野鳥音	環境音	バイク音	平均
Linear	単純	0.97	0.24	0.52	0.70	0.61
	MFCC	1.00	0.89	0.96	0.96	0.96
Poly.	単純	0.39	0.47	0.38	0.70	0.49
	MFCC	1.00	0.91	0.97	0.93	0.96
RBF	単純	0.96	0.59	0.59	0.70	0.71
	MFCC	1.00	0.79	0.96	0.96	0.93

4章における実験および本章での試みから、線形カーネルとRBFカーネルのどちらが有用であるかについてであるが、サンプル数が増した状態においても高い分類精度を示した線形カーネルを用いることが妥当では無いかと思われる。多項式カーネルは4クラスにおける分類において高い精度を出すことには成功したが、4章の精度の悪さに鑑みると使用を避けるのが無難であると思われる。また、サンプル数が300と、まだまだ多いとは言えない状況であるのも確かである。このことから、サンプル数が増えた場合においても線形カーネルを用いたほうが良いのかどうかという点については要検証事項である。特に、サンプル音源の個数が千や万単位などになった場合でも果たして線形カーネルで分類できるのかという疑念が抱かれる。

次章では本研究について考察やまとめを行い、また、本研究における技術的な課題や今後の可能性などに関して議論する。

第 6 章

結論

本章では、本研究により得られた結果や成果の総括、およびリモートセンシングにおける今後の可能性について述べてゆく。

6.1 本研究のまとめ

既存のリモートセンシングおよび実地調査における問題点は、研究背景において挙げた中川らによる動物を用いた無線伝送システム [5] が解決する可能性が高いと考えられる。しかし、その動物間通信には通信を行う動物が森林や山間部にいると GPS を介した位置情報の取得ができなかったり、タイムサーバに接続できないことから時刻補正ができなかったりする問題点が存在した。そこで、本研究ではそのような利用可能な情報源が限られる地帯においても定期的に利用することができるであろうと考えられる航空音を用いることにより、野生動物の位置情報を同定する手法を提案した。また、その手法を実現するにあたり必要と考えられる環境音から航空音を分類する手法の提案および実験から評価までを行った。なお、この際における環境音とは航空音、野鳥音、環境音、バイク音と 4 つの音要素を想定している。

提案した分類方法は機械学習と音声認識を用いた分類器の生成であり、機械学習では教師あり学習のサポートベクトルマシンを分類器として用いた。また、音声認識には高速フーリエ変換により得られた特徴量として用いた手法を単純処理、および人の聴覚特性を反映させ環境音分類に有効とされるメル周波数ケプストラム係数を特徴量として用いた MFCC 処理と定義した。そして、双方の音声処理と分類器においては三種類のカーネル (Linear, Polynomial, RBF) を用いた分類を行った。また、分類の際に用いた音源は、任意の 3 箇所 (船田, 栃本, 矢竹) における音源を実際に人間が聴き取り教師データを作成した。その際、船田の音源は東京大学サイバーフォレスト研究チームが提供する web サービスである Cyberforest からダウンロードした非可逆圧縮音源を用いた。しかし、船田の音源は非可逆圧縮音源より高周波帯の成分が切り落とされている点が懸念された。それより、栃本と矢竹は可逆圧縮音源で録音することにより、高周波帯において航空機音に特徴

的なスペクトラムが存在しないかどうかを判断する音源とした。この可逆圧縮音源を解析してみたところ、やはり航空機音における高周波帯には特徴的なスペクトラムが見受けられなかった。これより、先行研究には反するが、本研究では高周波帯において特徴的なスペクトラムが無いと仮定し実験を行った。

実験の際に用いた音源は上述した通りであり分類結果は 3 箇所×3 カーネル×2 処理方式の計 18 個の結果が得られることになる。分類を行った結果から、MFCC 処理の分類精度が高いことが判明した。また、分類器自体の性能を評価する尺度として ROC 曲線を用い評価を行ったところ、Linear および RBF カーネルの評価値が高いと言う結果を得ることができた。

しかし、航空音の教師データを作成するにあたり、矢竹の音源は記録時間が少なく、かつ、航空音自体の入力が少なく航空機の音が鳴り響いている間に、野鳥音がノイズとなる状況が数多く見受けられた。これより、我々は野鳥音を分離することにより精度を向上させられる可能性があるかと図り、航空音における特徴的なスペクトラムが見受けられる帯域以外をカットするフィルタ設計した。また、サンプル数が少ないことから分類精度が高く出ているのではないかと言う懸念から 3 箇所の音源を混合することにより分類器の精度を再評価した。この際に、矢竹では乗り物の入場が禁止されていた経緯からバイク音を含んだクラスを 4-Class と呼び、含まないクラスを 3-Class と定義した。それぞれの航空機音の音源に設計したフィルタに通し、改善案を用いた検証を行った。この検証より得られた結果から、さらなる分類精度の向上が判明した。カーネルごとの分類結果からは、フィルタを通さない手法では Linear と RBF、フィルタを通した手法では Linear と Polynomial カーネルの分類精度が高かった。なお、フィルタを通さない分類において Polynomial カーネルの分類精度は劣悪であったが、フィルタを通した後には向上したことより、Polynomial カーネルは音源次第で分類結果に大きな変化が起こる可能性があると思われる。これらを踏まえると、この程度のサンプル数においては Linear カーネルを用いることが妥当であると考えられる。しかし、サンプル数が増えた場合において同じ結果が得られるのかどうかと言う問題があり、その場合においては Linear カーネルよりも非線形である RBF カーネルを用いたほうがより良い結果を得る可能性があるかもしれない。その点に関しては、今後検証してゆく必要がある。

6.2 本研究の残された技術的な課題

前章にて挙げたように、航空音の音源にローパスフィルタを通すことにより、混在すると考えられる野鳥音を分離することで分類精度の向上を実現した。しかし、分類を行う際に用いるカーネル次第では結果が左右されることも判明した。特に、ローパスフィルタをかけた MFCC 処理における RBF カーネルを用いた結果は野鳥音の分類において AUC 値が低いことが判明した。これは RBF カーネルのような非線形なカーネルであると分離し難い

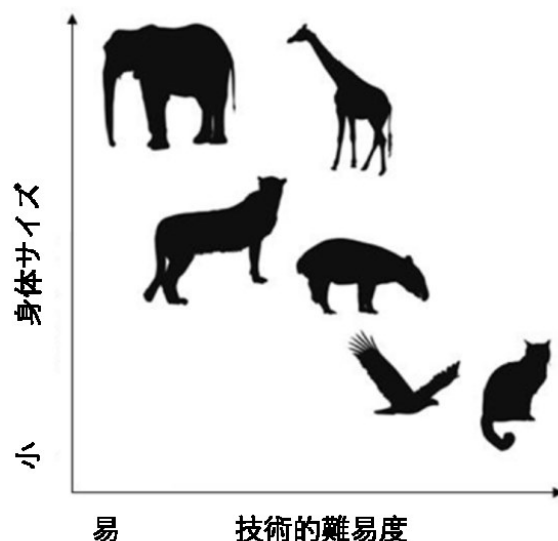


図 6.1 野生動物にセンシングデバイスを搭載するにあたる技術的難易度 (左) ¹³

超平面に特徴量がプロットされたからではないかと考えられる。また、サンプル数が数百と少ない現段階では線形カーネルによる分類が円滑に行われているだけで、この数が増加した場合には非線形カーネルの有用性が高くなる可能性も考えられる。この点に関しては、先述したように今後検証してゆく必要がある。

さらに、現段階では人間の耳を頼りに航空音かどうかと言う真値を判断しているが、その際には航空音だと認識でき、かつ音源編集ソフトの Audacity において -20db 以上の入力があり、かつ航空音の音源と人間が認識できる場合において航空音だと定義づけている。しかし、この定義方法だと今回使用した音源には存在しなかったような、定義した音圧以下で航空音が生じていた場合などにおいて誤分類を行う可能性がある。したがって、そのような状況においても問題が生じないような航空音の定義方法について懸念する必要があると思われる。

また、この環境音分類手法が実際に使用されることを想定すると、今回用いた固定マイクと比較し野生動物に搭載するデバイスに付属するマイクでは性能が劣ると考えられる。そこで、野生動物に大型で高性能なマイクを搭載すれば良いのではないと思われるが、動物に装着可能なシステム総重量は体重の 2%程度と定められている [19]。図 6.1 では野生動物にマイクを搭載するうえでの技術的難易度と身体サイズの関連性についてあらわした図であるが、小型な動物になればなるほどマイクを搭載する難易度が高くなることが伺える。サバンナのようなゾウやキリンと言った動物が存在するような区域におけるセンシングを行うのであればマイクの性能に対しての懸念はしなくても良いように考えられる。しかし、そのような動物が存在しないような区域のためにも、高性能でもない小型マイクに

¹³ 図は H. Kobayashi [19]より一部加工を施し引用

おける音源においても高い分類精度を得ることができるかどうかを今後検証してゆく必要があると思われる。

他の技術的な問題点として、現段階では限られた場所における晴天下の環境音において分類を実現することには成功している。しかし、これが雨天や強風の条件下では晴天時と同じような結果を得られるとは考えにくい。なぜなら、撮影用のマイクに雨粒が落下した際や強風がマイクに吹き込むことによるノイズが考えられるからである。雨天時に関しては、雨よけなどのカバーをすることにより回避することは考えられるが、最終的に、野生動物に搭載した小型マイクにおいては、野生動物は自然環境を走り回ることが想定されるため、これは避けがたい問題となろう。この問題に対してのアプローチも今後必要であると考えられる。

また、本研究では設計する際に用いたプログラム言語は Python とインタプリタ型言語であることより、プログラムの実行速度が懸案事項である。今回、実験章において明記はしていないが、実験で用いたサンプル数程度では実行速度が気になりになることは無かった。しかし、仮にサンプル数が膨大になった場合は分類器が学習する際の計算量が増大することが想定される。しかし、今後サンプル数が増えることによって、プログラムの実行速度に影響が出るのではないだろうかという懸念が生じている。今後、これらの事項を踏まえサンプル数が多くなった場合においても高速な演算を実現するために Python やそれに関連する演算ライブラリに依らないようなシステムを再構成することが必要になる可能性がある。

次に、今回の実験においては、非可逆圧縮音源である mp3 と可逆圧縮音源である wav の両音源を用い実験を行った。その際、双方の音源において高い分類精度を示せたことから音源の可逆・非可逆形式による分類精度の懸念はしなくても良いと考えられる。しかし、同じ mp3 でもビットレートやチャンネル数などに起因する音質の違いにより分類精度に差が生じる可能性は大いに考えられる。特に、野生動物に搭載したマイクから動物間通信を行うことにより最終的に基地局へ記録データを送り届ける際にはファイルサイズが大きいとファイル情報に誤りやそもそも搬送に失敗する可能性が考えられる。それより、ファイルサイズの小さい記録方式においても、高い精度で分類が可能であるのかどうかという点についても検証する必要がある。

これらの課題を解決することにより、今後、この技術の実用性が増すであろうことが言えよう。

6.3 今後の可能性

6.1 節で挙げたような課題を検証および解決してゆくことや、自動で野生動物の位置同定および時刻補正を実現することにより、中川らが提案した動物を用いた新たなリモートセンシング手法の一環として、この技術が貢献できるのでは無いかと考えられる。それを

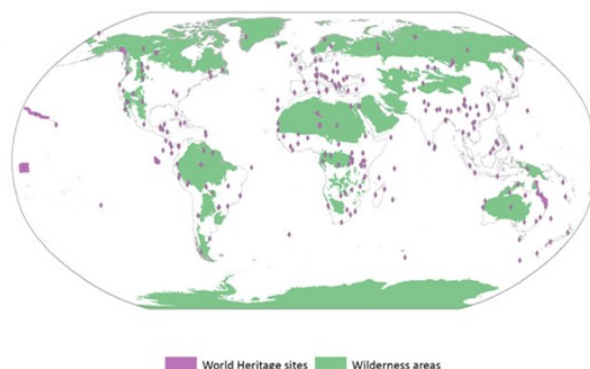


図 6.2 全世界における自然保護区域 (右)¹⁴



図 6.3 アオウミガメを用いた海洋環境の空間情報センシング機構の概要図¹⁵

実現してゆくためにも、今後、自動でデバイスを搭載した野生動物の位置を同定するシステムの完成が急がれる。そのシステムの完成後には、この技術と動物間通信を用いた伝送システムを共用することにより、実際にリモートセンシングを行いその結果を評価する必要がある。

この手法が確立することにより、これまでリモートセンシングを行うことが難しいと考えられていた地帯におけるセンシングが、野生動物さえ存在していれば可能となるとが想定される。また、このリモートセンシング手法が確立された後には、大気汚染や空中線量など任意のセンサを用いることにより得られた値が自然環境などに与える影響を定量的に把握することが可能となるであろう。

Kormos によると全世界における陸地の 4 割が自然保護区域 (Wilderness Areas) であると述べている [31]。ここで言う自然保護区域とは、人間が自然に対して介入を行っていないような地帯である。図 6.2 では、その全世界における自然保護区域 (図上での緑色の箇所) を示す。このような条件下における地域では、そもそも人間が長期間に亘る実地調査を行うことは研究背景でも述べたように、立ち入りが制限されたような区域への制限解除

¹⁴ 図は C.F.Kormos ら [31]より一部加工を施し引用

¹⁵ 図は工藤ら [32]より引用

の承認をうけるまでに時間がかかることや、それに伴う人的コストなどがかかることから難しく、そのような場合に当センシング手法が有効になると言えよう。さらに、工藤らはこのセンシング手法をアオウミガメに適用することにより、海洋環境の生態系や海中における線量などをセンシングするような技術を考案している [32]。具体的には、アオウミガメには単独行動と集団行動を行う習性があり、単独行動の時はデータの収集、集団行動の際にはデータの共有と、これらの習性を活用することによりリモートセンシングが行えることを想定している。また、帰巣本能を利用することによりデータ収集用のデバイスを充電することも可能であるとしているとしている。図 6.3 はその概要を示している。つまり、センシングエリアを陸だけではなく海にまで広げようとする試みが行われているのである。しかし、工藤らでの例では電波の入る、いわば時刻補正のかけられるような区域での使用が考えられている。例えば、無人島における浜辺においてウミガメが長期に亘り電波が入らないケースも考えられる。その際に、本研究および動物間通信の技術が合わさることにより、時刻補正をかけることが可能になると考えられる。これより、陸上のみならず海中においてセンシングエリアを拡充させるようなリモートセンシング手法を実現するための技術の一端として貢献できると思われる。

この動物間通信を用いたセンシング手法は動物が主役であることから、センシングをするにあたって特別な通信網を通すような工事をしたり、センシング結果をモニタリングするような環境を整えたり、電力を引いてくる必要がない。必要な物は、CPPS において作成された専用の端末と携帯や固定回線による無線通信環境だけである。原生林や原発事故などにより立ち入りが禁止されているような地帯ともなれば、そもそも存在していない、またはそれまでに存在していた生活インフラが無くなっていることが考えられる。原生林に関しては、センシングを行う際にそれらの設備を整えるのだけでも相当の年月を費やす必要がかかるであろう。また、原発事故などの現場では施工を請ける企業を探すことや、そもそもその施工を国から認可してもらう必要がある。このように従来手法では多大な手間が必要と考えられるが、今までどおりに原生林や原発事故以前の環境において生活をしてきた動物にデバイスを装着するのみであり、手間がかからなくなることは非常に有利な点であることが言えよう。また、このセンシング手法は原発事故現場周辺などと長期間の滞在が難しいと言われる場所においてもセンシングを行うことができる。そのセンシングによって得られた結果が最終的には社会に還元をもたらすであろうことを期待する。

謝辞

本論文は、東京大学大学院 新領域創成科学研究科 社会文化環境学専攻 修士課程に在籍中に、筆者が取り組んだ研究をまとめたものです。

はじめに、大学院へ入学する以前から進学の相談や本研究に対しての内容や方向性に対してのご指導のみならず、学生生活や進路など研究以外における点でも様々な知識をご教授していただき、多岐にわたる面においてお世話になった瀬崎薫 教授に絶大なる感謝を申し上げます。心よりお礼申し上げます。

また、本研究を行うにあたり研究テーマとして提示していただき、研究を行う上で数多くのご指導や問題点に対する解決策の議論などと大変お世話になった小林博樹 講師に多大なる感謝の意を表明します。秩父演習林での実験および郡山における調査や、解析の際に用いる機材の支援などがあったことより、これら実験や解析作業を円滑に行うことができました。また、国内外における学会発表に向けた論文執筆やそれに伴う現地発表用資料に対してご指導いただいたこともあり、結果として非常に貴重な経験をさせていただきました。改めて、感謝の意を表明いたします。

本研究に対しての様々なアドバイスや執筆論文に対する助言をはじめ、数多くのご指導をいただいた伊藤昌毅 助教に深く感謝いたします。伊藤助教とは、これまでの情報技術からこれからのテクノロジーはどうなってゆくのかと言った内容について話し合ったことが印象的であり有意義なものでありました。ありがとうございました。

東京大学 空間情報科学研究センター 特任研究員の中村和彦 氏には秩父における音源収録の際には演習林における宿舎の手配やそれに関わる事前準備などを手がけていただいたこともあり、現地での収録を円滑に行うことができたうえ、システムを設計してゆくうえで具体的にどのようなことを考えればよいのかと言ったイメージを膨らませることができました。また、東京大学 空間情報科学研究センター 学術支援員および東京藝術大学 社会連携センター 特任研究員の佐々木遊太 氏とは秩父での実験を行った際に、研究内容に対してのディスカッションや異なる分野の人から見た物事の捉え方などこちらも新たな知見を得ることができました。こちら両名に対しまして、感謝いたします。

東京大学 瀬崎研究室の OB である、清水和人 氏、木田裕一朗 氏、中川慶次郎 氏、坂本敬太 氏、牛コウ 氏、加藤弘則 氏、稲葉瞳 氏、中村直人 氏、森英記 氏、牧山紘 氏、山本直人 氏、鵜飼祐太 氏、松野有弥 氏、Dunstan Matekenya 氏には、研究の他に論文の執筆や学生生活や進路についてなど様々な情報をいただきました。瀬崎研究室の博士課

程である青木俊介氏（現在，Carnegie Mellon University に所属），江甜甜氏，孫堯氏，中山悠氏とは，研究内容やどのような分野の研究が今盛んに行われているのかと言った内容について議論することができました．また，修士課程であり同期である星野光玖氏，鈴木孝男氏，西井香織氏とは研究や，進路についての相談と楽しい時間を過ごすことができました．そして，後輩である北里知也氏，伊藤匡一氏，梅沢啓佑氏，Ben Rukutantichoke氏，楊珂為氏からは研究に対する熱意やこれまで研究室に無かったものを導入していこうという姿勢に対して刺激を受けました．また，学会参加やその他における数多くの事務手続きを行っていただいた秘書の松本夏穂氏，内藤潤氏に感謝いたします．

また，本研究室に進学するきっかけや相談をはじめ後押しをしていただいた，青山学院大学 理工学部 戸辺義人 教授にはこの場において感謝の意を表明いたします．

この他にも，数多くの方々のご厚意があったからこそ，本論文を執筆することができたことは言うまでもありません．本当にありがとうございました．

最後に，18年間と言う歳月に亘り，何不自由なく有意義な学生生活を送らせてくれた両親と様々な価値観を与えてくれた妹に心からの感謝を表明することで，本論文をしめさせていただきます．

2017年1月23日 合間 優陽

発表文献

- Y. Kamma, K. Sezaki, H. Kobayashi, "Spatio-Temporal Information Correction Mechanism for Wild Animal Wearable Sensors", 1st Asian Students Symposium on Emerging Technologies (ASSET 2016), Singapore, June 2016.
- 合間 優陽, 大西 鮎美, 中村 和彦, 瀬崎 薫, 小林 博樹, "野生動物装着センサ用の空間情報補正機構の基礎的検討", マルチメディア, 分散, 協調とモバイル (DICOMO2016)シンポジウム, 3E-3, 2016年7月
- 合間優陽, 瀬崎薫, 小林博樹, "立ち入りが制限された地域における航空音自動認識を用いた位置推定手法の提案", 電子情報通信学会ソサイエティ大会, 知的環境とセンサネットワーク, B-18-32, 2016年9月

参考文献

- [1] 避難指示区域の状況 - ふくしま復興ステーション - 福島県ホームページ (オンライン), <http://www.pref.fukushima.lg.jp/site/portal/list271-840.html>, 最終閲覧日: 2017-01-05.
- [2] 伊藤 昌毅, 中川 慶次郎, 小林 博樹, 瀬崎 薫, "動物を利用した環境調査のための遭遇検知による効率的なデータ転送手法の提案," マルチメディア, 分散, 協調とモバイル (DICOMO2014) シンポジウム, pp.605-612, 2014 年 7 月.
- [3] 小林 博樹, 松島 潤, "イリオモテヤマネコおびきよせ&タッチアンドゴー機構-野生動物装着型センサの空間情報回収方法の研究," 2012 年東京大学空間情報科学研究センターシンポジウム (CSIS DAYS 2012), 2012 年 11 月 2-3 日.
- [4] 大島 浩太, 利根川 秀, "近距離・長距離通信の連携による野生動物行動検知のための無線センサネットワークの提案," マルチメディア, 分散, 協調とモバイル (DICOMO2015) シンポジウム, 4D-4, 2015 年 7 月.
- [5] K. Nakagawa, H. Kobayashi, K. Sezaki, "Carrier pigeon-like sensing system: animal-computer interface design for opportunistic data exchange interaction for a wildlife monitoring application.," Proceedings of the 5th Augmented Human International Conference, pp. 12. ACM, Kobe, Japan, (2014).
- [6] 林野庁/都道府県別森林率・人工林率 (平成 24 年 3 月 31 日現在) (オンライン), <http://www.rinya.maff.go.jp/j/keikaku/genkyou/h24/1.html>, 最終閲覧日: 2017-01-05.
- [7] 野外安全マニュアル - 日本生態学会 (オンライン), <http://www.esj.ne.jp/safety/manual/>, 最終閲覧日: 2015-01-05.
- [8] 警視庁生活安全局地域課, "平成 27 年における山岳遭難の概況," (2017) .
- [9] 全国地質調査業協会連合会 労働災害検討チーム, "地質調査業務における労働災害の現状と対処," (2015) .
- [10] 国土交通省 関東地方整備局 企画部 技術調査課, "平成 28 年度 工事事故防止「重点的安全対策」," (2016) .
- [11] 東京大学工学系等安全衛生管理室. 安全衛生管理業務支援システム-事故・災害事例一覧 (オンライン), <http://anzen2.t.u-tokyo.ac.jp/anzen2/>, 最終閲覧日: 2017-01-20.

- [12] L. Kaleschke, G. Heygster G. Spreen, "Sea ice remote sensing using AMSR-E 89-GHz channels," JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 113, 2008.
- [13] 産総研：ランドサット 8 号の日本上空からのデータを即時公開（オンライン），
http://www.aist.go.jp/aist_j/new_research/2013/nr20131122/nr20131122.html, 最終閲覧日：2017-01-07.
- [14] "INES(国際原子力・放射線事象評価尺度)評価," 環境省, 2015.
- [15] "International Nuclear and Radiological Event Scale(INES)," International Atomic Energy Agency, 1990.
- [16] K. Ohno, S. Kawatsuma, T. Okada, E. Takeuchi, K. Higashi, S. Tadokoro, "Robotic control vehicle for measuring radiation in Fukushima Daiichi Nuclear Power Plant," 2011 IEEE International Symposium on Safety, Security, and Rescue Robotics, 2011.
- [17] H. Cruickshank, S. Farrell, M. Marchese C. Caini, "Delay- and Disruption-Tolerant Networking (DTN): An Alternative Solution for Future Satellite Networking Applications," Proceedings of the IEEE, Vol. 99, 2011.
- [18] K. Makiyama, K. Nakagawa, M. Katayama, M. Nagasawa, K. Sezaki, H. Kobayashi, "Synchronization of Peripheral Vision and Wearable Sensors for Animal-to-Animal Interaction.," The 16th International Conference on Human-Computer Interaction (HCI 2015), 2015-08.
- [19] H. Kobayashi, "Basic Research in Human-Computer-Biosphere Interaction.," The University of Tokyo, 2010.
- [20] Flightradar24.com - Live flight tracker!（オンライン），
<https://www.flightradar24.com/>, 最終閲覧日：2015-01-08.
- [21] B. Bertrand, R. Christophe, M. Jean-Mariec, "A pattern recognition approach for aircraft noise detection," Internoise, 2009.
- [22] M. Takagi, K. Fujimoto, Y. Kawahara, T. Asami, "Detecting Hybrid and Electric Vehicles Using a Smartphone," UbiComp, 2014.
- [23] Y. Rui, A. Gupta, A. Acero, "Automatically Extracting Highlights for TV Baseball Programs," MULTIMEDIA '00 Proceedings of the eighth ACM international conference on Multimedia, pp. 105-115, 2000.
- [24] Z. Liu, Y. Wang, T. Chen, "Audio Feature Extraction and Analysis for Scene Segmentation and Classification," Journal of VLSI Signal Processing Systems - special issue on multimedia signal processing, vol. 20, pp. 61-79, 1998.

- [25] 山田 侑太朗, 西村 雅史, 峰野 博史, "個人差のない特徴量を用いた水分摂取量推定手法の検討," 情報処理学会マルチメディア, 分散, 協調とモバイル(DICOMO2016)シンポジウム, pp.1814-1820, 2016.7.
- [26] C.M. ビショップ, 元田 浩, 栗田 多喜夫, 樋口 知之, 松本 裕治, 村田 昇, パターン認識と機械学習 下: 丸善出版, 2012-04-05.
- [27] D. W. Robinson, R. S Dadson, "A re-determination of the equal-loudness relations for pure tones," British Journal of Applied Physics, vol.7, 1956.
- [28] W. Han, C. Chan, C. Choy, K. Pun, "An Efficient MFCC Extraction Method in Speech Recognition," 2006 IEEE International Symposium on Circuits and Systems, pp.4, 2006-05.
- [29] 3.3. Model evaluation: quantifying the quality of predictions — scikit-learn 0.18.1 documentation (オンライン),
http://scikit-learn.org/stable/modules/model_evaluation.html#accuracy-score, 最終閲覧日 : 2017- 01-15.
- [30] Bing Tian Dai. scikit-learn/base.py (オンライン),
<https://github.com/scikit-learn/scikit-learn/blob/14031f6/sklearn/svm/base.py#L610>, 最終閲覧日 : 2017- 01-19.
- [31] C.F. Kormos, B. Bertzky, T. Jaeger, Y. Shi, T. Badman, J.A. Hilty, B.G. Mackey, R. A. Mittermeier, H. Locke, E. Osipova, J.E.M. Watson, "A Wilderness Approach under the World Heritage Convention," A journal of the Society for Conservation Biology, 2015-01.
- [32] 工藤 宏美, 小林博樹, "アカウミガメを用いた海洋環境の空間情報センシング機構," CSIS DAYS 2013, 2013-01.
- [33] サービスエリア | エリア | NTT ドコモ (オンライン),
<https://www.nttdocomo.co.jp/support/area/>, 最終閲覧日 : 2017- 01-05.
- [34] Sensor types (オンライン),
http://www.eorc.jaxa.jp/hatoyama/experience/rm_kiso/mecha_sensortype.html, 最終閲覧日 : 2017- 01-07.