

## 3. 音声解析を用いた猛禽類の生息判定技術

### 3-1 音声解析を用いた生息判定技術の開発の背景と目的

これまで猛禽類調査では、調査員による林内踏査や直接観察によって、営巣場所の特定や繁殖状況の把握に多大な労力を要してきた。一方、限られた時間内での調査では、発見漏れの危険もある。これを改善し、効率的に現地調査を行うことを目的として、簡易に猛禽類の生息判定を行うための技術開発を試みることにした。

3章では、1章（表 1-3-1）で整理した技術の中から、鳴き声録音（音声解析）技術の開発に取り組む。

鳴き声録音（音声解析）の対象種は、希少猛禽類の中でも、道路事業において最も頻繁に調査対象となるオオタカを選んだ。

オオタカが頻繁に調査対象となる理由は、オオタカが日本中に分布し、かつ都市近郊の林や人里近い里山に生息するためである。また、他の希少猛禽類に比べれば生息数が多いため、事業地周辺に生息する可能性も高いことが挙げられる。

他方、鳥の鳴き方は、種類によって特徴（周波数や継続時間）が異なり、同じ種内でも様々な鳴き方のパターンが存在する。

そこで今回は、オオタカについて先行的に音声解析を用いた生息判定技術の開発を試みた。



## 3-2 音声解析を用いたオオタカの生息判定技術の開発

### (1) オオタカの生息判定技術の開発のための作業フロー

今回開発を試みたオオタカの生息判定技術は、調査対象地周辺で録音した音声データから、オオタカの音声データを機械的に抽出し、生息の有無を簡易判定する第1段階「スクリーニング」と、抽出した音声をオオタカの鳴き声のパターン（警戒、餌乞、交尾、ヒナ、幼鳥）と比較し、統計的に自動分類する第2段階「音声判別モデル」から構成されている。

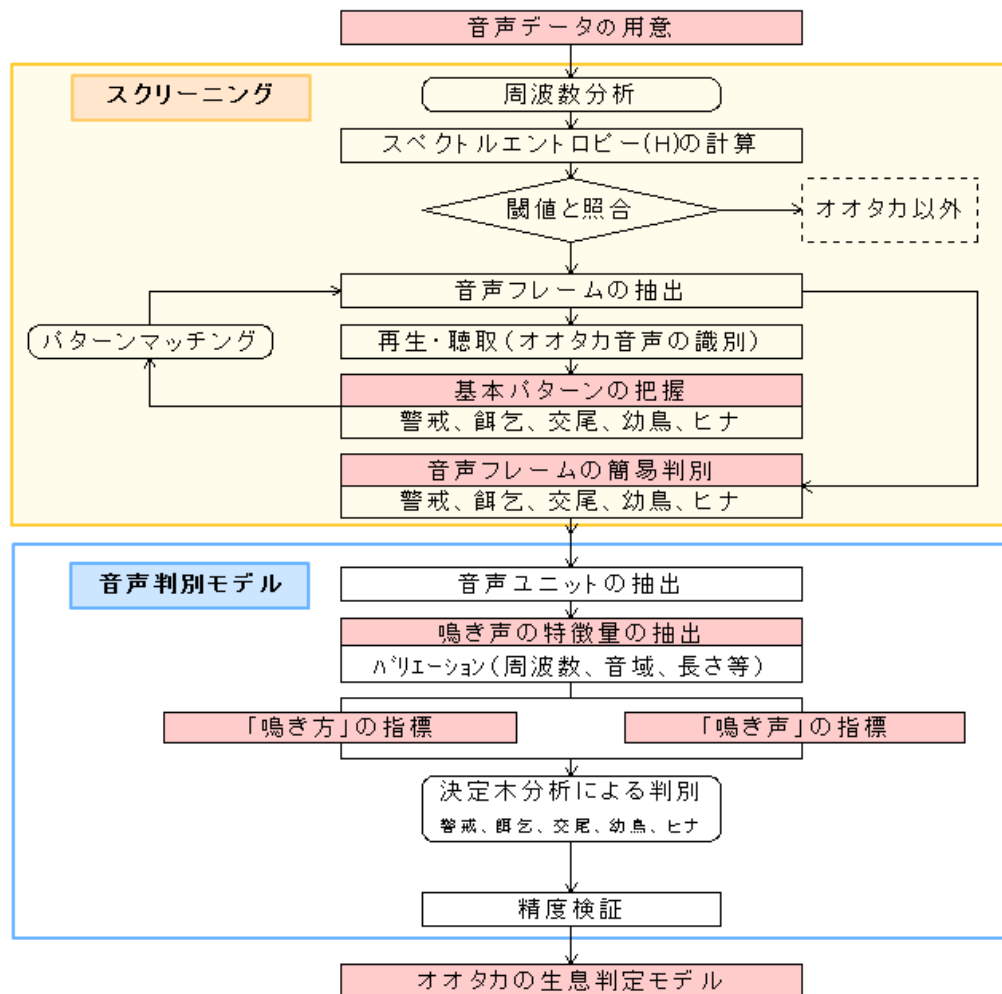


図 3-2-1 音声解析によるオオタカの生息判定技術の開発のための作業フロー

## (2) オオタカの生息判定技術の開発方法

### 1) 音声データの用意

オオタカの音声判別モデルの構築に使用する音声データは、関東地方で録画されたオオタカの巣内観察ビデオ（A地区、B地区、C地区の3地区）を基に、映像から抽出した音声を用いることとした（写真3-2-1、表3-2-1）。これら音声データの抽出には、一般的なフリーソフトを用い、WAV形式、無圧縮、フレームレート48000Hzまたは8000Hzで抽出した（表3-2-2）。なお音声は、カメラ付属のマイクロホンにより、採録されている。

撮影は、各地区とも繁殖期間中の午前4時～午後7時の間に行われていたが、音声解析にはオオタカの活動時間帯である午前8～12時の4時間のデータを用いた。

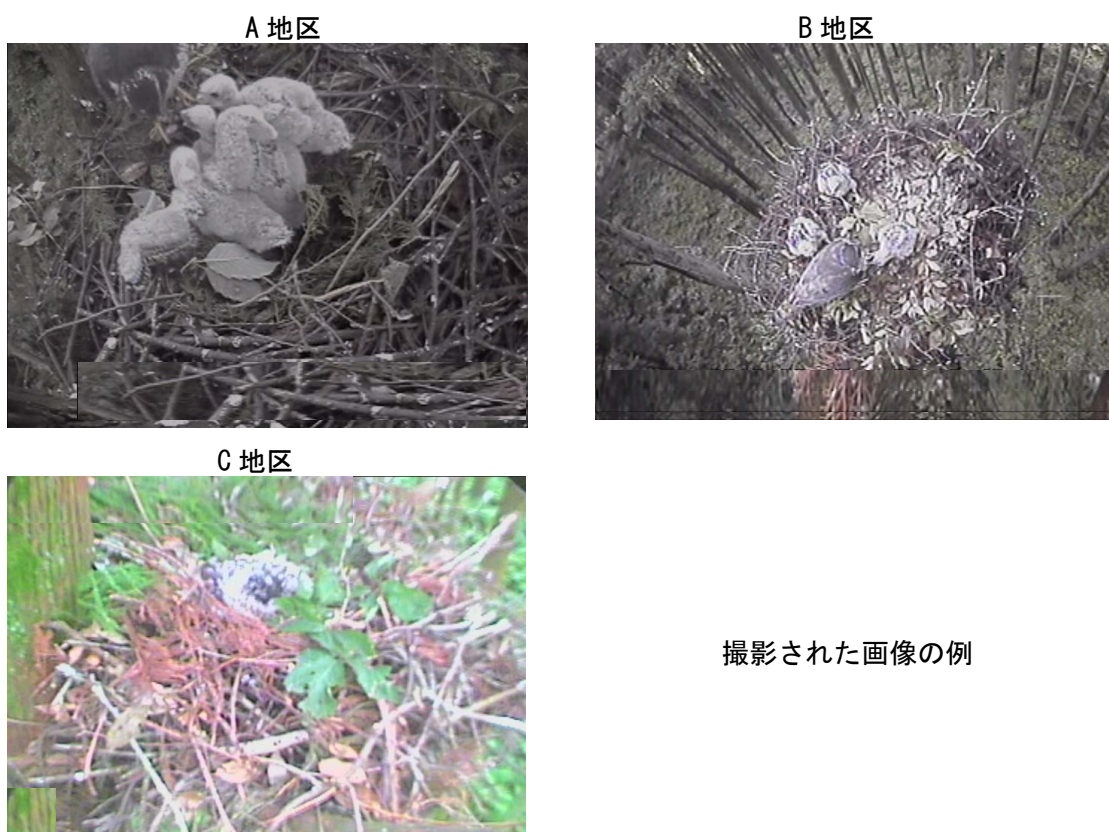


写真3-2-1 巣内ビデオによる各巣の録画状況

表3-2-1 巣内ビデオの撮影年と撮影期間（音声データの取得状況）

地区\年	2008	2009	2010	2011	2012	2013	2014	のべ時間
A地区	2-7月	2-8月	2-7月		4-8月			2542時間
B地区			3-8月	2-7月			3-7月	1988時間
C地区			4-7月	1-7月			3-7月	1694時間

表 3-2-2 巢内ビデオから抽出した音声データの諸元

諸元		A 地区	B 地区	C 地区
映像	圧縮形式	MPEG2	MPEG2	独自
	解像度 (ピクセル)	352×240	352×240	704×480
	フレームレート(FPS)	29	29	2
音声	圧縮形式	MP2	MP2	ADPCM
	サンプリング周波数(Hz)	48000	48000	8000
	ビットレート(bps)	224000	224000	32000
	チャンネル数	2	2	1

## 2) スクリーニング

ここでは、調査対象地周辺で録音した音声データから、オオタカの音声を機械的に抽出し、生息の有無を簡易判定するための手法について検討した。

### 1. スペクトルエントロピーの計算と音声フレームの抽出

巢内ビデオから抽出された音声には、オオタカの鳴き声以外にも、様々な環境音やノイズが含まれている。また音声解析では、連続した音声データは扱えないため、これらを短い区間に区切り、個々に解析の対象とする必要がある（これら短い区間に区切ったものを、以下「音声フレーム」という）。

そこで、周波数分析によって得られるスペクトルエントロピー（式(1)）を指標として、音声フレームを抽出することとした。

なおスペクトルエントロピーは、ホワイトノイズ等の均一な信号では高い値となり、有意な音声等の不均一な信号では低い値となる。そのためスペクトルエントロピーを用いることで、一定の周波数を示す環境音やノイズを除去することができ、かつオオタカの鳴き声のように1回あたりの継続時間がまちまちな音声を抽出することが可能になる。

抽出作業では、全ての音声データについてスペクトルエントロピーを計算し、あらかじめ設定した閾値（前後1分間の平均値）と比較し、閾値以上のスペクトルエントロピーを示した音声区間を「音声フレーム」として抽出した（表3-2-3）。

$$H = - \sum_{k=M}^N p_k \log_2 p_k, \quad p_k = \frac{s_k}{\sum_{i=M}^N s_i} \quad \dots (1)$$

ここで、 $H$  : スペクトルエントロピー  
 $p_k$  :  $k$  番目の周波数のパワーの存在確率  
 $s_k$  :  $k$  番目の周波数のパワー、  
 $M, N$  : 周波数の下限値、上限値

表 3-2-3 抽出できた音声フレームの数

地区\年	2008	2009	2010	2011	2012	2013	2014
A 地区	913,850	1,093,043	779,422		1,077,727		
B 地区			915,192	882,586			1,135,138
C 地区			415,870	880,692			601,648

## 2. 基本パターンの把握

オオタカの鳴き声には、バリエーションがある。

そこで、抽出した音声フレームの一部を聴取し、鳴き声を「警戒」「餌乞」「交尾」「幼鳥」「ヒナ」の5種類に分類した（表 3-2-4）。

このうち、ヒナと幼鳥は、前者は巣内育雛期前半（孵化から 24 日目ごろまでの羽毛に覆われた時期）、後者は巣内育雛期後半（25 日目以降から巣立ちまで）として区分した（写真 3-2-2）。

表 3-2-4 オオタカの鳴き声の基本パターン

種別	鳴き声のパターン	備考
警戒	ケッケッケツ	主に成鳥のオス
餌乞	クァイクァイー	主に成鳥のメス
交尾	クェークェークェー（オス） キャツキャツキャツ（メス）	オス、メスの鳴き声が重なり合って聞こえる
幼鳥	キッキッキッフィアー	警戒や餌乞に近いが、より高音
ヒナ	クィークィー/ピー ピー	鳴く機会は少ない

### 巣内育雛期前半（ヒナ）



### 巣内育雛期後半（幼鳥）



写真 3-2-2 巣内育雛期前半および後半の状況

### 3. パターンマッチング

オオタカの音声判別モデルの構築には、モデル構築用のデータ（学習データ）と構築したモデルを検証するためのデータ（検証用データ）が、それぞれ必要である。

しかし元の音声データは、長時間かつオオタカ以外の音声が含まれるため、これを、逐一、人が聴取し、学習データと検証用データを用意するのは現実的ではない。

そこで、前述の2. において鳴き声を基に5つの基本パターンに分類済みの「音声フレーム」を用いて（表3-2-5）、パターンマッチングを行うこととした。

パターンマッチングでは、5種類の基本パターンそれぞれについて、鳴き声の周波数分布（図3-2-2～3）を基に、未聴取・未分類の「音声フレーム」からパターンの似た音声を機械的に抽出した。これを人が聴取して、鳴き声の種別を判定し、その中から代表的なものを、新たなパターンマッチングの指標（教師データ）として使用した。これらパターンマッチングを繰り返し、オオタカの音声が入った多くの「音声フレーム」を得た（表3-2-6）。

なお、パターンマッチングは、Dynamic Programming マッチング法（系列になったデータ同士の類似度を比較する方法）により行った。また、判別の精度を高めるため、音量（音の振幅）の変動係数（標準偏差／平均値）が1未満の音声フレームはパターンマッチングの対象から除外した。

表 3-2-5 パターンマッチングに使用した音声フレーム数

種別	A 地区	B 地区	C 地区
成鳥	30	74	42
餌乞	16	142	99
交尾	8	18	23
幼鳥	29	230	632
ヒナ	14	23	55



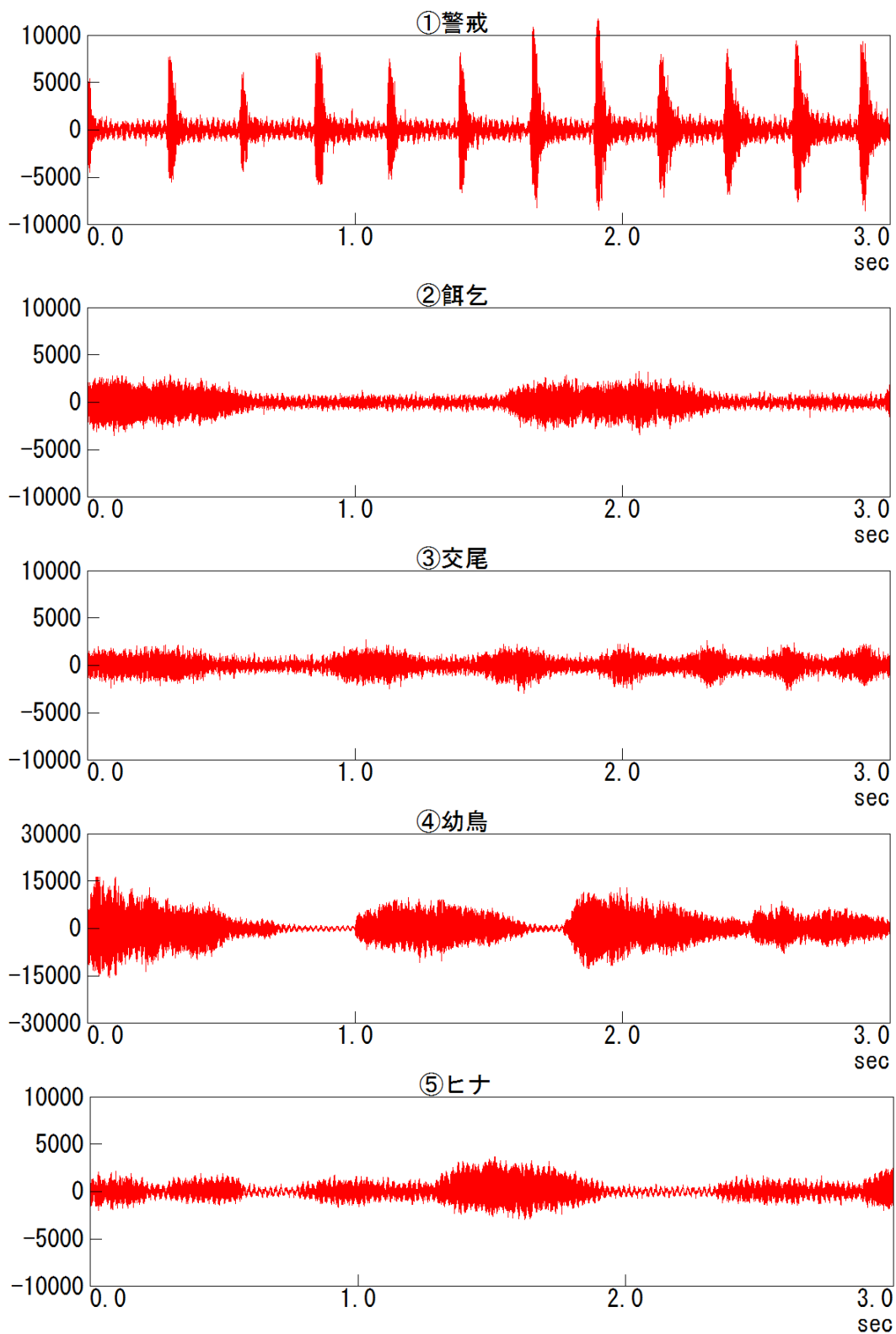


図 3-2-2 オオタカ音声の波形の例  
縦軸は音の大きさ



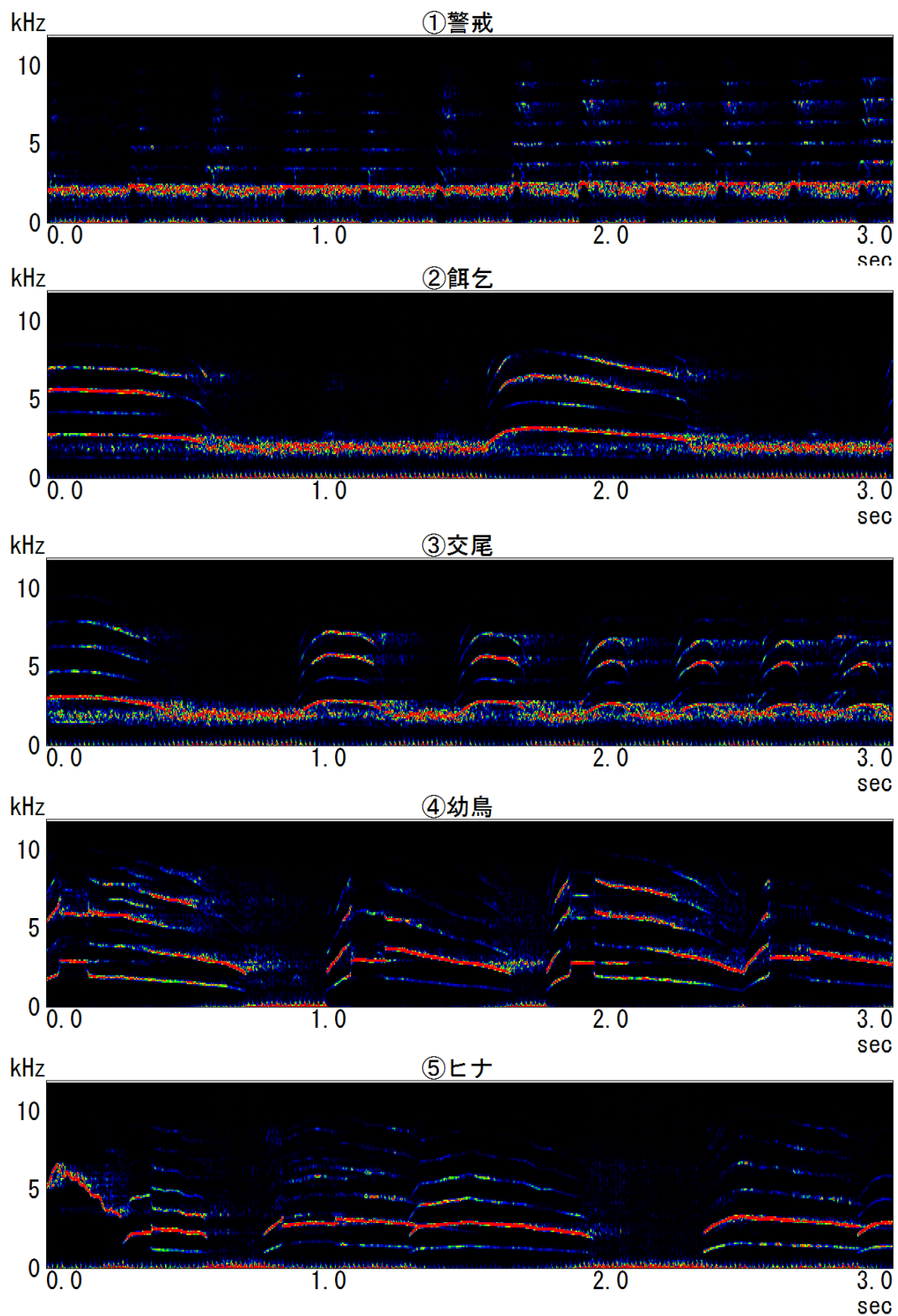


図 3-2-3 オオタカ音声のスペクトログラム（声紋）の例  
 図中の色は青から赤になるほど音が強いことを示している

表 3-2-6(1) パターンマッチングによる音声フレームの簡易判別結果 (A 地区)

年	月	警戒	餌乞	交尾	幼鳥	ヒナ	計
2008	3		2				2
	4	2,064	157	1,171	1		3,393
	5			25	10		35
	6				3		3
	7				5	7	12
	計	2,064	159	1,196	19	7	3,445
2009	3		3	6			9
	4	1,872	483	976	2		3,333
	5	35	1	67	1		104
	6				176	8	184
	7				47	3	50
	計	1,907	487	1,049	226	11	3,680
2010	3		4	11			15
	4	871	30	727			1,628
	5	324		87	3		414
	6		1		27	2	30
	7				9	1	10
	計	1,195	35	825	39	3	2,097
2012	4		2	116	2,969	1	3,088
	5		4	25	1,267	9	1,305
	6		13	7	1,812	32	1,864
	7			3	11,563	24	11,590
	8				22,609	6,825	29,434
	9			1	5,068	4,250	9,319
	計	0	19	152	45,288	11,141	56,600

表 3-2-6(2) パターンマッチングによる音声フレームの簡易判別結果 (B 地区)

年	月	警戒	餌乞	交尾	幼鳥	ヒナ	計
2010	2	1	2				3
	3		151	6	23		180
	4		290	2	165	1	458
	5		73	3	290	1	367
	6		180	13	1,687	17	1,897
	7		57	5	579	1	642
	計	1	753	29	2,744	20	3,547
2011	1		2				2
	2		37	1	59		97
	3		74	4	13		91
	4		149		123		272
	5		66	2	158	1	227
	6		294	13	1,226	2	1,535
	7		95	2	330		427
	計	0	717	22	1,909	3	2,651
2014	3		18		2		20
	4		49		31	1	81
	5		12	1	48		61
	6		44		125	2	171
	7						0
	8						0
	計	0	123	1	206	3	333

表 3-2-6(3) パターンマッチングによる音声フレームの簡易判別結果 (C 地区)

年	月	警戒	餌乞	交尾	幼鳥	ヒナ	計
2010	3		1			3	4
	4	1	299		1	88	389
	5	2	1,451		179	1,118	2,750
	6	1	1,076		27	3,117	4,221
	7		376		39	2,803	3,218
	8		11			52	63
	計	4	3,214	0	246	7,181	10,645
2011	3		1				1
	4		3				3
	5		15			21	36
	6		69			152	221
	7		4		1	293	298
	8		2			34	36
	計	0	94	0	1	500	595
2014	3		2				2
	4						0
	5					4	4
	6		10		5	258	273
	7		2		11	440	453
	計	0	14	0	16	702	732

### 3) 音声判別モデル

ここでは、スクリーニングによって抽出した音声を、5種類のオオタカの鳴き声のパターン（警戒、餌乞、交尾、ヒナ、幼鳥）に統計的に自動分類する「音声判別モデル」を構築する。

#### 1. 音声ユニットの抽出

音声フレームには有音と無音の区間があるが、ここでは、有音の区間を「音声ユニット」として定義した。

オオタカの鳴き声などの音声は、有音と無音の繰り返しで表現できる（図3-2-4）。そこで、音声フレームに含まれる有音区間（音声ユニット）の数と割合を、次式(2)、(3)で求めた。ここでは、音声フレーム1つあたりの音声ユニットの数を「音声ユニット数」、1音声フレームに占める音声ユニットの長さを「音声ユニット率」と定義した。

$$R = \frac{\sum_{i=1}^m \Delta t_i}{T} \quad \dots (2) \quad , \quad N = \frac{m}{T} \quad \dots (3)$$

ここで、 $R$ : 音声ユニット率、 $N$ : 音声ユニット数、 $T$ : 音声フレーム時間（秒）  
 $\Delta t_i$ :  $i$ 番目の音声ユニット時間（秒）、 $m$ : 音声フレーム内の音声ユニット

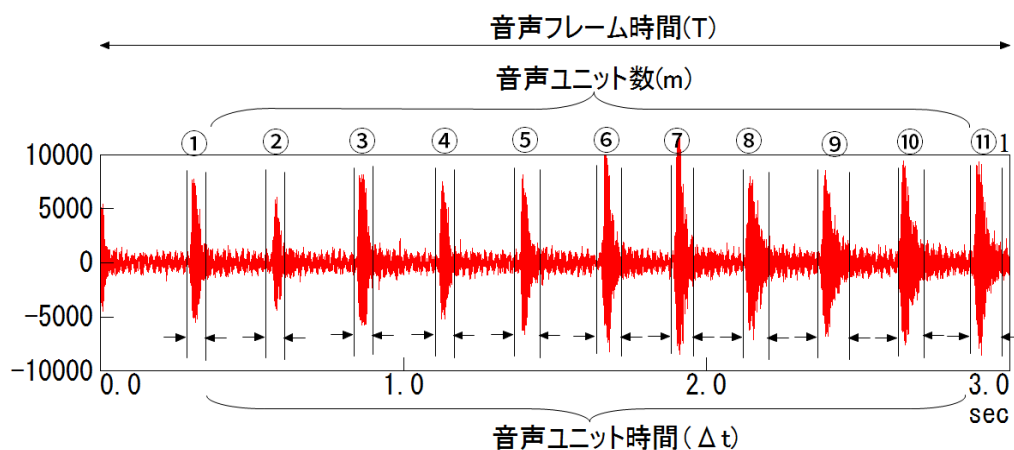


図 3-2-4 鳴き声の特徴（鳴き方の指標）

縦軸は音の大きさ、横軸は時間

## 2. 鳴き声の特徴量の抽出

オオタカの鳴き声の特徴を表す指標には、1. で調べた有音と無音の繰り返しで表現できる「音声ユニット数」、「音声ユニット率」だけでなく、鳴き声の周波数などの特徴量がある。

そこで、各音声フレームの周波数分布上でのピークを求め、ピークの大きな順に3つの周波数を、鳴き声の特徴量として採用した（ただし1kHz以下のピーク周波数は雑音として除いた）。

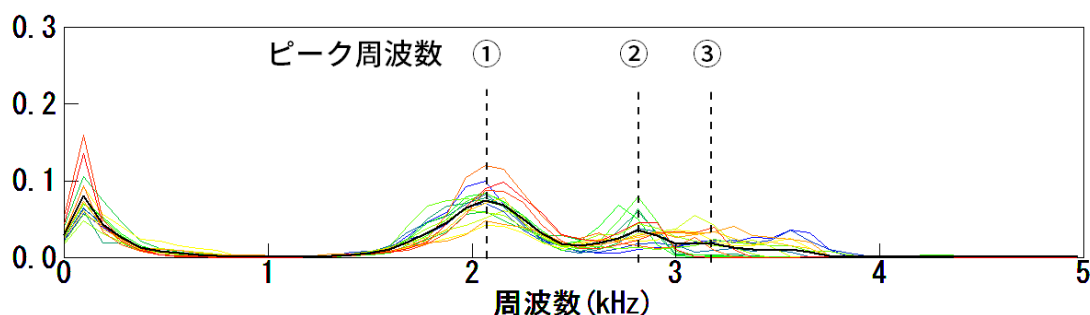


図 3-2-5 モデルの構築に用いる鳴き声の特徴（「鳴き声」の指標）

縦軸は相対強度（周波数全体を1とした時の値）、横軸は周波数 (kHz)

① ～③はピーク強度が上位3つまでの周波数

## 3. 鳴き方と鳴き声の指標

オオタカの鳴き方の特徴を表す指標として、1. で調べた有音と無音の繰り返しで表現できる「音声ユニット数」と「音声ユニット率」を解析に用いることとした。

オオタカの鳴き声の特徴を表す指標として周波数分布、すなわち2. で調べた各音声フレームの上位3つのピーク周波数の値を用いることとした。なお、ピーク周波数の値は音声フレームごとに絶対値の大きさが異なるので、周波数全体を1とした時の相対強度で補正した値を用いた。

#### 4. 決定木分析

機械学習の一種である決定木分析を行い、5種類のオオタカの鳴き声のパターン（警戒、餌乞、交尾、ヒナ、幼鳥）に統計的に自動分類するための閾値を計算した。なお決定木分析とは、結果が分岐図で表される判別分析の一種である。

決定木分析には、あらかじめ人が聴取し、鳴き声のパターンを分類済みの音声フレーム（表 3-2-5）を解析に用いた。この分類済みの音声フレームの半数を、モデルを構築するために使用する学習データとし、残り半数を構築したモデルの精度検証に用いる検証データとした。決定木分析の目的変数には、この学習データを用い、判別を行うための説明変数には、「鳴き声」の指標であるピーク周波数（周波数全体を1とした時の相対強度）と「鳴き方」の指標である音声ユニット数と音声ユニット率を用いた。

これらの判別モデルは地区別に構築した。

決定木分析の結果、A地区では、まず、最大ピークの周波数（V1）で判別され、次に音声ユニット数（V5）と音声ユニット率（V4）で判別された（図 3-2-7(1)）。B地区、C地区では、いずれも最初の判別は音声ユニット数（V4）であり、次いで最大ピークの周波数（V1）と音声ユニット数（V5）であった（図 3-2-7(2)～(3)）。

これらの判別項目とその閾値の組み合わせが、オオタカの鳴き声を判別するための基準、すなわち「オオタカの音声判別」モデルである。

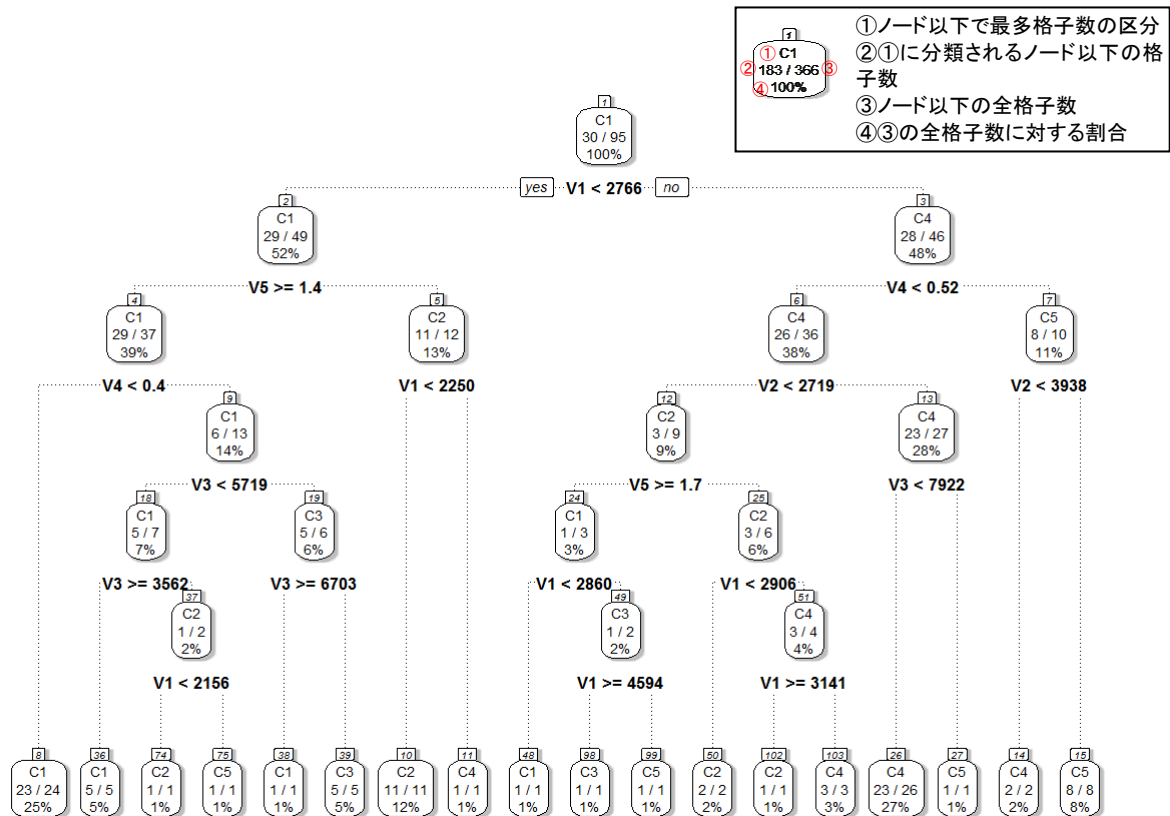
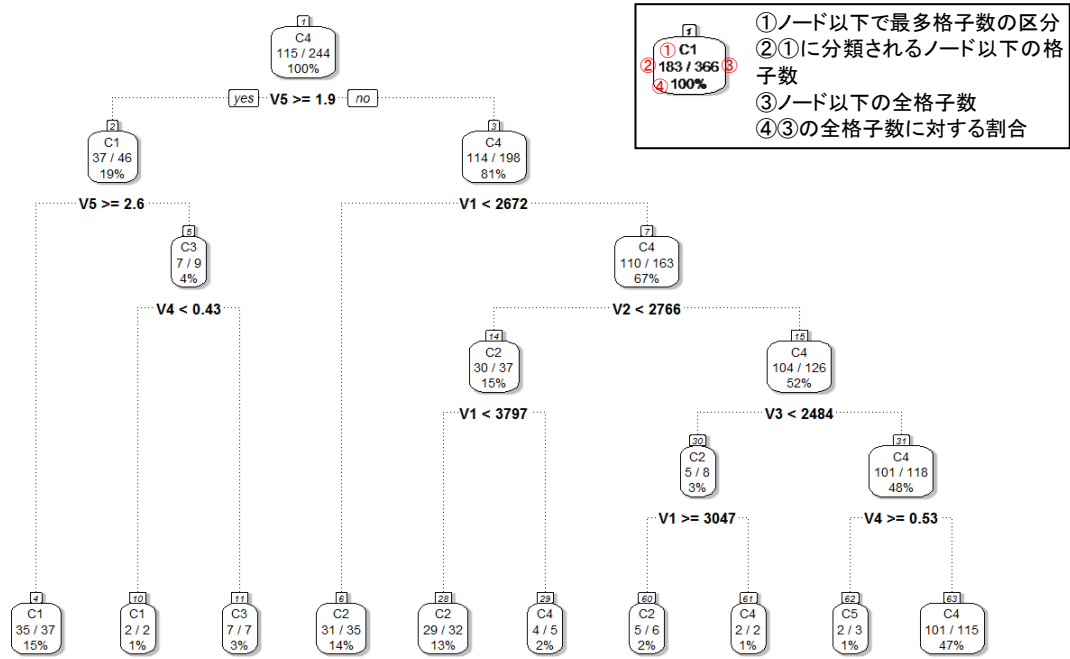


図 3-2-6(1) オオタカ音声判別モデル (A地区)

目的変数	C1: 成鳥 (警戒) C2: 成鳥 (餌乞) C3: 成鳥 (交尾) C4: 幼鳥 C5: ヒナ
説明変数	鳴き声: V1: ピーク周波数 (1位)、V2: ピーク周波数 (2位)、V3: ピーク周波数 (3位)
	鳴き方: V4: 音声ユニット率: $\Sigma$ 音声ユニット時間 / 音声フレーム時間
	V5: 音声ユニット数: $\Sigma$ 音声ユニット数 / 音声ユニット時間



①ノード以下で最多格子数の区分  
 ②①に分類されるノード以下の格子数  
 ③ノード以下の全格子数  
 ④③の全格子数に対する割合

図 3-2-6 (2) オオタカ音声判別モデル (B 地区)

目的変数	C1: 成鳥 (警戒) C2: 成鳥 (餌乞) C3: 成鳥 (交尾) C4: 幼鳥 C5: ヒナ
説明変数	鳴き声 V1: ピーク周波数 (1位)、V2: ピーク周波数 (2位)、V3: ピーク周波数 (3位)
	鳴き方 V4: 音声ユニット率: $\sum$ 音声ユニット時間 / 音声フレーム時間 V5: 音声ユニット数: $\sum$ 音声ユニット数 / 音声ユニット時間

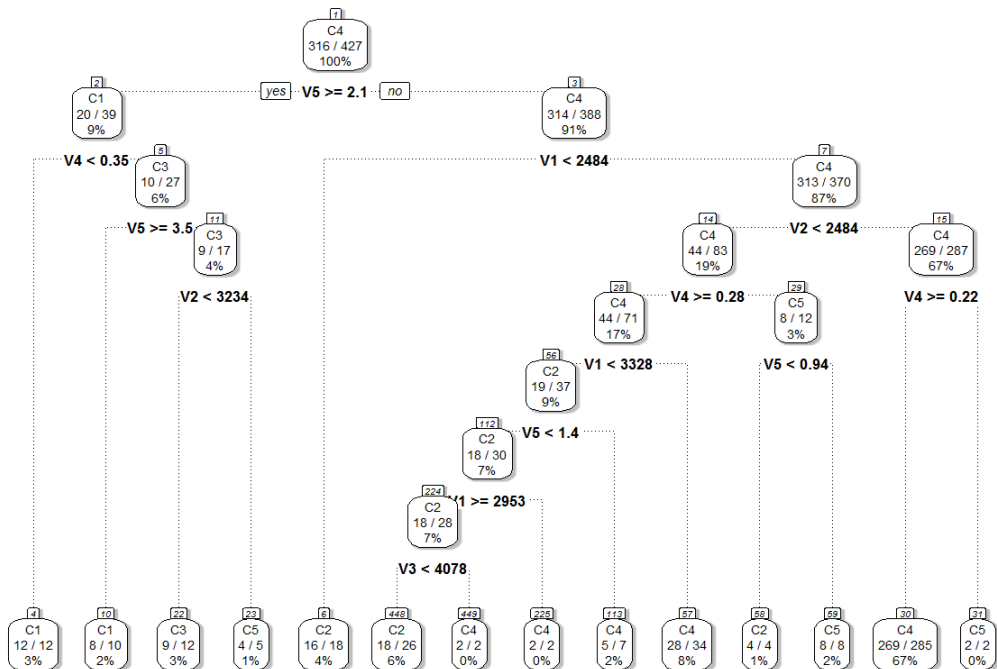


図 3-2-6 (3) オオタカ音声判別モデル (C 地区)

目的変数	C1: 成鳥 (警戒) C2: 成鳥 (餌乞) C3: 成鳥 (交尾) C4: 幼鳥 C5: ヒナ
説明変数	鳴き声 V1: ピーク周波数 (1位)、V2: ピーク周波数 (2位)、V3: ピーク周波数 (3位)
	鳴き方 V4: 音声ユニット率: $\sum$ 音声ユニット時間 / 音声フレーム時間 V5: 音声ユニット数: $\sum$ 音声ユニット数 / 音声ユニット時間



## 4) 音声判別モデルの検証

### 1. モデルの検証方法

決定木分析によって構築したモデルの妥当性を、分類結果の適合率（分類結果の正答率）を用いて、2通りの組み合わせで評価した。

まず、同一地区内での適合率を調べた。これは、あらかじめデータを学習データと検証データの2つに分けておき、一方の学習データのみを用いて構築したモデルを使って、もう一方の検証データの判別を行い、その適合率を調べる方法である。また学習データのみを用いて構築したモデルを使って、あらためて構築使用した学習データの判別を行い、その適合率を調べる方法もある。

なお適合率の計算は、数式(4)で表される。

(当該地区の検証データによる評価)

$$r = \frac{m^*}{m} \quad \dots \quad (4)$$

ここで、 $r$  : 適合率

$m^*$  : モデルの分類と種別が一致した数

$m$  : 検証データの数

次に、地区間での適合率を調べた。これは、ある地区で構築した判別モデルを、他地区のデータ（学習データと検証データからなる全データ）の判別を行い、その適合率を調べる方法である。

なお適合率の計算は、数式(5)で表される。

(他地区の学習データによる評価)

$$r' = \frac{M^*}{M} \quad \dots \quad (5)$$

ここで、 $r'$  : 適合率

$M^*$  : モデルの分類と種別が一致した数

$M$  : 他地区の学習データの数

## 2. モデルの検証結果

同一地区内で、学習データのみを用いて構築したモデルを使って、学習データおよび検証データの判別を行った結果、同一地区内では 80 %台後半の高い適合率を示した。

表 3-2-7 同一地区での音声判別モデルの検証結果（適合率）

種別 \ 地区	A 地区	B 地区	C 地区
学習データ	94.7 %	86.9 %	86.2 %
検証データ	—	87.7 %	87.5 %

A 地区については、データ数が少なかつたため、検証データによる検証は行っていない

ある地区で構築した判別モデルを使って、他地区のデータ（学習データと検証データからなる全データ）の判別を行った結果、約 65 ~71 %の適合率を示した。

表 3-2-8 地区間での音声判別モデルの検証結果（適合率）

モデル \ データ	A 地区	B 地区	C 地区
A 地区	—	68.4 %	71.6 %
B 地区	65.2 %	—	65.2 %
C 地区	68.6 %	68.3 %	—

本研究で示された地区間で 65%以上という適合率は、録音条件（マイクの周波数特性）が異なり、かつ巢内ビデオに含まれる圧縮処理された音声を用いたことを考えれば、比較的高い適合率であると考えられる。

したがって、本モデルに改良を加えることで、様々な地区に対応可能な判別モデルが構築できる可能性がある。

### 3-3 オオタカの生息判定技術の活用場面と活用上の留意点

今回、環境アセスメント等において調査事例の多いオオタカを対象に、音声解析を用いて簡易に生息判定を行う技術「オオタカの生息判定技術」の開発を試みた。

このオオタカの生息判定技術は、調査対象地周辺で録音した音声データから、オオタカの音声を機械的に抽出し、生息の有無を判定する第1段階と、抽出した音声をオオタカの鳴き声のパターン（警戒、餌乞、交尾、ヒナ、幼鳥）と比較し、統計的に自動分類する第2段階から構成されている。

本技術は、関東地方の3つの巣内ビデオの音源を用いて試作したものであり、野外での検証は不十分である。今後、データを蓄積し、判別の精度を向上させることで、環境アセスメント等でのオオタカ調査の効率化が図られることだろう。

具体的には、下記のような調査の効率化方法が考えられる。

- 1) 猛禽類の生息有無を、長期間の録音記録から把握することができる。オオタカの生息が予想される林に、ICレコーダー等の録音機器を数日から1週間程度設置する。その後、ICレコーダーから回収した録音データを本技術によって解析することで、オオタカの鳴き声を自動で抽出し、生息の確認を容易に行うことができる。
- 2) あるいは、オオタカの繁殖状況を知ることができる。オオタカの営巣木の近傍に録音機器を設置し、録音データを本技術によって解析することで、鳴き声のパターンから、巣での行動内容（警戒、餌乞、交尾）や、雛の成長段階（卵の孵化日や雛の巣立ち時期）を自動的に判別することができる。

本技術は、試作段階であり、野外での検証が不十分である。

そのため、使用にあたっては、以下の点に留意いただきたい。

- ・ 解析で設定した指標や決定木分析で示された閾値は、試行段階のものである。そのため、抽出漏れ・誤判別の可能性が残されている。また、録音条件等によって判別精度が異なる可能性がある。
- ・ あくまでオオタカ用であり、鳴き声の特徴（周波数など）が異なる種には適用できない。
- ・ 音声解析は、オオタカが生息していないことを確認（不在を証明）するためのものではなく、オオタカが生息している証拠を効率的に発見するための補助的な技術である。

