

G_001

時刻ごとの音量最大周波数を基準とした野鳥の品種識別

Classification of Bird Species Based on Frequency with the Maximum Power at Each Time

牧野 洋平† 三田 長久† 東谷 幸治†
 Youhei Makino Nagahisa Mita Kouji Higashitani

1. まえがき

科学技術の応用により自然環境保護を目指す観点から、自然環境保護評価システムの開発を目指し、その一端として、自然環境の変化に敏感な野鳥の調査を目的とし、音声情報を用いた野鳥の品種識別法についての提案を行う。

本研究では複数のフォルマントを用いた識別法^[1]をもとにするが、小さな音声にも対応できるようにスペクトログラムを作成し、その分影響が大きくなる雑音への対策としてフォルマントを最も大きいもの一つ、すなわち時刻ごとに音量が最大となる周波数のみを用いての識別を目指す。具体的にはまず音声信号に短時間フーリエ変換を行ない、スペクトログラムで表す。計算量削減の為、時間軸方向の識別範囲を限定し、そこから識別に用いる値(素音)を抜き出す。抜き出した素音があらかじめ用意しておいたテンプレートと周波数がどれだけ異なるか、周波数の変化の仕方がどれだけ異なるかを数値化して識別を行う。

以下、第2章で時間軸の限定と素音の抽出方法、第3章で素音とテンプレートの比較方法、第4章で音声データを用いたシミュレーション結果、第5章で総括としての結論を述べる。

2. 特徴の抽出

2.1 短時間フーリエ変換

本研究では予め用意しておいた長さ1秒の音声で識別を行った。まず識別にかける音声に短時間フーリエ変換を行ない(式1)時間を横軸、周波数を縦軸、音量 $(|ST(t, f)|^2)$ を0~255に正規化し白黒の濃淡で示すスペクトログラムを作る。窓関数はハニング窓を使用した。ここで時間の刻み幅は $T=1/220(s)$ とした。

$$ST(t, f) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)e^{-j2\pi f\tau} d\tau \quad (1)$$

$$h(t) = 0.5 - 0.5 \cos\left(\frac{2\pi t}{T}\right) \quad (0 \leq t < T) \quad (2)$$

この際に1秒の中で最大の $|ST(t, f)|^2$ が255となるように正規化することで識別する音声小さくても識別を可能にする(図1)。

2.2 時間方向の制限

1秒の中には殆ど情報が含まれない時間帯が存在するので、使用する時間の制限を行う。

まず224以下の音量を雑音として削除し、時刻ごとに残った音量の合計を取る。合計が最大となった時刻を中心に取り、識別を行うのに十分な時間として30/220秒を識別にかける。

2.3 素音の抽出

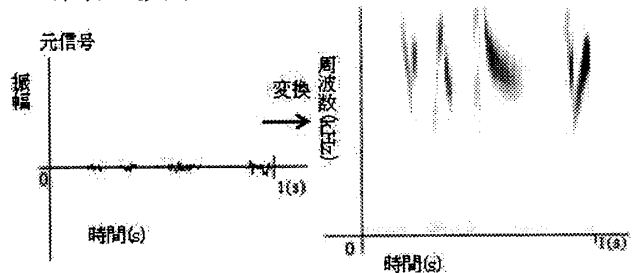


図1 微小な信号の短時間フーリエ変換

そこから時刻ごとに音量が最大の周波数を点として抽出す。具体的には音量が大きいものから順に30/220(s)の半分15/220(s)分から音量最大周波数の抽出を行う。全時刻から抽出を行わないのは2.2項での処理を行ってもまだ情報が含まれない時間帯が残っているためである。

そうして得られた音量最大周波数の集合を時間方向に不連続、周波数方向に大きく離れている点で分割し、識別にかけるのに十分な時間的長さがあるものを素音とする。そのため一つの音声から複数の素音を得られる場合もあり、その場合は全ての素音を識別にかける。素音とは本来は音声の塊の区分の一つであるが、本研究ではこのような音量最大周波数の集合のことを表す。

予め用意しておく識別する品種ごとのテンプレートはサンプル音声から得られる素音の中からその品種の特徴をよく表しているものを極力小さく選ぶ。例えば、サンプル素音が凸型ならば上端の尖っている部分、右下がりならば端の部分を除いて出現頻度の高い周波数範囲のみをテンプレートとする。今回は1品種ごとに2個ずつウグイス・キビタキ・ホオジロ・メジロ・ホトトギスについてテンプレートを用意した。

3. 識別方法

こうして抽出した素音と予め用意しておいたテンプレートが周波数方向にどの程度離れているか、周波数の変化の様子がどの程度異なるかについて比較する。以下、この二項目を識別距離 d 、識別形状 d_{hg} とする。そして識別距離と識別形状より最終的な識別値 d_{judge} を求める。

3.1 識別距離

† 熊本大学 大学院自然科学研究科

図2は素音とテンプレートとを□と○の集合で表したものである。識別距離は双方から基準となる時刻を選び(図2斜線部)、互いの基準時刻が等しくなるように時間軸方向に平行移動させ、基準時刻における双方の周波数差を取るものとする。

基準時刻は次項の識別形状が計算可能なように素音とテンプレートが重なる幅 W が $W > 1$ となるような全ての組み合わせを試し、最終的な識別値が最小になる組を選ぶものとする。

3.2 識別形状

識別形状は識別距離を0となるように、各時刻における周波数差 d_i の平均を取り

$$d_{fig} = \frac{\sum_{i=1}^{W-1} |d_i - d|}{W-1} \quad (3)$$

とする。ただし、基準時刻においては周波数差が0なので明らかなので予め計算から除外し、その分、素音とテンプレートが重なる時間 W を1減らす。

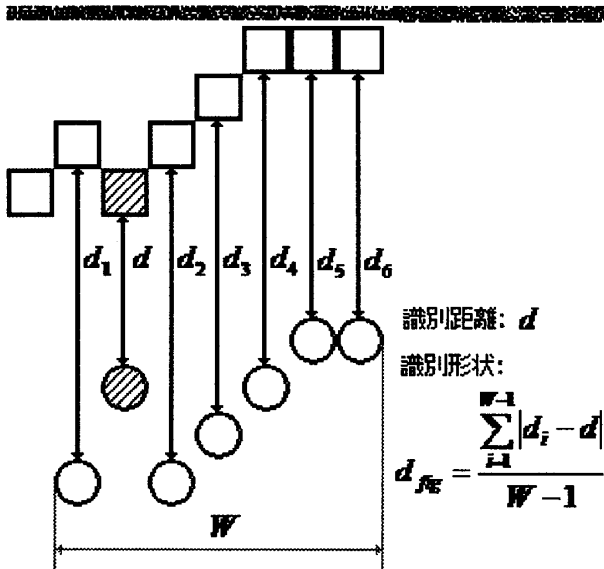


図2 識別距離・識別形状の定義

3.3 識別値

識別距離と識別形状を合わせて最終的な識別値を求める。識別距離は同一品種における鳴き声の高低の個体差を表す要素があるのでやや重要性が低い。その点を考慮して、識別形状が最小になるように基準時刻を定め、識別形状・識別距離を求めると、両者の最大値は概ね1:2となった、その比率より最終的な識別値を

$$d_{judg} = \frac{d}{2} + d_{fig} \quad (4)$$

とする。以降、基準時刻を識別値が最小になるように取る。

全テンプレートについて識別値を求め、最も識別値の小さい品種を識別結果とする。但し、識別距離・識別形状・識別値それぞれ個体ごとに上限値を設け(表1)どれか一つでも上限値を超えたらその値は無効とし、全品種に対して値が

無効となった場合、その音声は識別対象品種のものではないとする。

表1 各値の上限値(1pt=25Hz)

	ウグ イス	キビタ キ	ホオジ ロ	メジ ロ	ホトト ギス
識別距離	12	20	10	10	14
識別形状	10	6	6	10	5
識別値	10	11	8	7	10

4. 実験

識別実験にはテンプレートを用意した5品種についてテンプレートの作成に用いたサンプル音声と概ね同じような鳴き方をしている音声を1品種当たり5個体、1個体当たり1秒の音声を5つ用意し識別を行った。個々の音声について識別結果を求め、個体ごとでは過半数以上(3個以上)の音声で得られた結果を採用するものとし、その結果を表2に示す。

表2 識別試験の正答率(%)

	ウグ イス	キビ タキ	ホオ ジロ	メジ ロ	ホト トギ ス	全体
個々の 音声	80.0	88.0	64.0	64.0	80.0	75.2
個体	100.0	100.0	80.0	100.0	100.0	96.0

5. 結論

個々の音声についての識別精度は75.2%と改善が求められるが、個体ごとで考えると96.0%と高い精度が得られた。

識別失敗の傾向としては、ホオジロとメジロの精度が特に低い、これは双方のテンプレートが互いに似ており識別が困難だということもあるが、双方の音声から得られる素音の周波数が他の3品種に比べ広範囲に広がっていることも考えられる。テンプレート同士を比較した場合、ホオジロ・メジロは他の3品種と比べ全体的に周波数が高く区別が容易である。しかし、素音は余分な要素が多くなるため素音の基準時刻に周波数が低くなる時刻が選ばれた場合、比較的周波数の低い他のテンプレートに対しても識別距離が小さくなりやすく、素音の抜出しの際に如何にして余計な要素を省き、周波数の範囲を狭めるかが問題になる。

識別精度には現れていないが、テンプレートを用意していない品種の除外も問題となる。テンプレートごとに識別を行い最も識別値の小さい品種を識別結果とする識別方法では、どれか一つでも各値が上限値以下で識別値が求められる品種があれば、識別対象外にならず、テンプレートを用意していない品種の音声を識別にかけると誤識別を起こす可能性が極端に高くなる。

以上、今後の課題としては抜出す特徴の簡略化と、識別対象外の品種の除外が上げられる。

本研究は一部を環境技術開発等推進費の援助を受けて実施しました。

参考文献

1. 高木 旬也:「音声情報を用いたサウンドスペクトログラム及びフォルマントによる野鳥識別法」, 熊本大学大学院修士論文, 2002