

鳥類図鑑Hyperbookにおける鳴き真似を用いた検索方式

矢川雄一 田淵仁浩 村岡洋一

早稲田大学理工学部

本稿では、マルチメディアデータベースシステムである鳥類図鑑Hyperbookにおける鳴き真似を用いた検索方式を提案する。鳥の鳴き声を聞いた利用者はその鳴き声の真似を入力し、検索する。利用者は鳴き真似において、(1) 鳴き声を構成する音の数、(2) 鳴き声のリズム、(3) 鳴き声の大きさと高さ、を表現する。検索アルゴリズムではこれらの類似構造を反映するために、音の大きさと高さのパターンに着目し、鳴き声と鳴き真似の距離を各パターンのなす面積の差と定義する。1040個の鳴き真似を用いた評価によると、提案した類似構造を利用者が表現しやすい鳴き声で、50%以上の検索率を得ている。

A method for retrieval by user's imitation of birdcalls in Hyperbook on birds

Yuichi Yagawa Masahiro Tabuchi Yoichi Muraoka

School of Science and Engineering, Waseda University
3-4-1, Ohkubo, Shinjuku, Tokyo, 169 JAPAN

Hyperbook on birds is a multimedia electronic book system. The purpose of this paper is to propose a method for retrieval by user's imitation of birdcalls in Hyperbook on birds. A user who have heard a birdcall queries the bird by imitating the birdcall. By imitating, users express (1)the number of syllable in birdcalls, (2)the rhythm in birdcalls, (3)the stress and the intonation in birdcalls. The distance between a birdcall and a user's imitation is defined as the difference in the sound power contour and the fundamental frequency contour, which mean stress time rhythm and syllable time rhythm. The evaluation shows that more than 50% of users' retrieval succeeded.

1. はじめに

マルチメディアデータベースシステムでは、システムから利用者にマルチメディアデータを提示する機能だけでなく、利用者が対話の手段として積極的にメディアを利用できる仕組みが重要である。マルチメディアデータベースの応用として現在開発中の鳥類図鑑Hyperbookでは、利用者が聞いた鳥の鳴き声情報を利用して検索する手段として、口笛や口真似などの鳴き真似による検索を用意している。

一方、鳥の鳴き声と利用者の鳴き真似は、音響的に決して一致しない。ところが実際の利用者の鳴き真似を聞いてみると、「この鳥ではなかるうか」と推定できる場合がある。これは鳥の鳴き声と利用者の鳴き真似が類似しているからである。そこで鳥類図鑑Hyperbookでは、鳴き声と鳴き真似の類似度を定義し、検索を利用者の鳴き真似に類似した鳥の集合を取り出す操作としている。

鳴き声と鳴き真似の音響分析を行なうと、両者は平均振幅の包絡線と基本周波数パターンで類似している。そこで本稿では、平均振幅の包絡線と基本周波数パターンはともに連続したパターンと考え、DPマッチングを用いた比較方法を提案し、その評価を示す。

以下では、まず鳥類図鑑Hyperbookとその1つの検索機能である鳴き真似による検索について述べる。次に鳴き声と鳴き真似の類似点を指摘し、その類似点に着目した検索アルゴリズムを説明する。最後に提案した検索アルゴリズムを実験により評価する。

2. 鳥類図鑑Hyperbookと鳴き真似による検索

電子化図書Hyperbookシステム[1]は、音、絵、映像などの各種メディアを統合し、しかも積極的に利用することにより、従来の本を超えた使いやすさを利用者に提供することを目指したマルチメディアデータベースシステムである。現在、我々はその応用として鳥類図鑑Hyperbookを開発中である。本節では、まずその鳥類図鑑Hyperbookについて説明する。さらに鳥類図鑑Hyperbookにおける鳴き声検索として鳴き真似による検索について説明し、その実現のための課題を指摘する。

2.1 鳥類図鑑Hyperbook

一般に市販されている鳥類図鑑によると、鳥を検索する手掛かりとなる情報には、シルエット、色のパターン、鳴き声、大きさ、場所と季節、行動がある[2]。鳥類図鑑Hyperbookでは、各情報毎にカスタ

マイズされた検索機能が用意されており、利用者は手掛かりとなる情報を入力し、検索する。ところが入力された情報は、元の鳥の情報と決して一致することがない。そこで鳥類図鑑Hyperbookでは、入力された情報に最も適合するもの、すなわち最も類似するものから順に利用者に提示する。さらに利用者は提示された鳥の姿を眺めたり、鳴き声を聞いたりして確認する。最終的に自分の求める鳥を探すことができた利用者は、その鳥について新しい知識を得ることができ、また獲得した知識は次回からの検索にも利用できる[3]。

2.2 鳴き真似による検索とその課題

2.1で述べたように、鳴き声は鳥を検索する際の手掛かりとなる情報である。特に鳥においては、姿は見えぬが鳴き声が聞こえたという状況がしばしばある。そこで鳥類図鑑Hyperbookでは、利用者が聞いた鳥の鳴き声のイメージを用いて検索する方法を用意し、これを鳴き声検索と呼んでいる。

これまで鳴き声検索として、鳴き真似による検索と印象による検索を提案している[4]。鳴き真似による検索では、利用者は鳥の鳴き声イメージを口真似、口笛などの鳴き真似で表現する。印象による検索では、利用者は鳥の鳴き声イメージを形容詞などの音の印象を現す言葉により表現する。

本稿で述べる鳴き真似による検索について考えると、利用者の鳴き真似は鳥の鳴き声と音響的に決して一致しない。そこで従来から提案している距離空間データモデルMeSOD[1]を採用し、鳴き真似による検索を、利用者の鳴き真似と似ている鳴き声の集合を取り出す操作としている。鳥類図鑑Hyperbookは入力された利用者の鳴き真似と鳴き声との類似度を計算し、図1のように近いものから順に利用者に

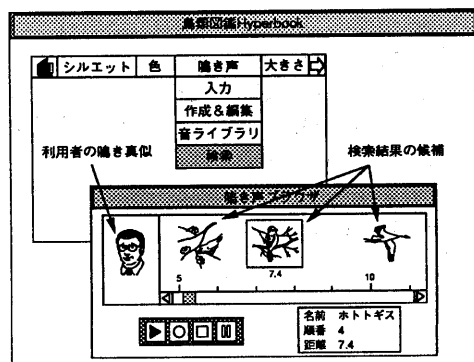


図1 鳴き真似による検索

提示する。そして利用者は自分がイメージし、表現した鳴き真似と最も近い鳴き声から順に確認する。

実際の利用者の鳴き真似を聞いてみると複数の鳴き真似で、直感的に「この鳥ではなかろうか」と推定できる場合がある。この際、鳴き真似には鳴き声の何らかの情報が表現されており、人はその情報を元に鳴き声を推定している。鳥類図鑑Hyperbookで、この人による検索と同様な検索機能を実現するためには、以下の2点が必要である。すなわち(1)人が鳴き声の何を真似ているかを明らかにすること(鳴き真似の類似構造の解明)、(2)明らかになった類似構造に基づいた類似度の計算アルゴリズム(距離の定義)、である。

3. 鳴き真似の類似構造

本節では、鳥の鳴き声と利用者の鳴き真似について2種類の音響分析を用いて比較した結果、明らかになった鳴き真似の類似構造について説明する。

3.1 鳴き声と鳴き真似の音響分析

一般に音の心理属性として、音の大きさ(強さ)、高さ、音色が考えられる[5]が、ここでは音の物理属性と比較的明確に対応のとれる音の大きさと高さについて考える。音の大きさは物理的には音圧波形の平均振幅に対応し、音の高さは音の基本周波数に対応する。鳴き真似は鳴き声の大きさと高さの相対的な変化の様子を表現していると仮定し、平均振幅の包絡線、基本周波数パターンについて調べた。

・平均振幅の包絡線

平均振幅の包絡線は、音声波形の8msecごとに32msec区間の平均振幅を計算し、最大値で正規化して求めた。図2の(a)はコノハズクの鳴き声の平均振幅の包絡線であり、(b)と(c)は2人の被験者による鳴き真似の平均振幅の包絡線である。

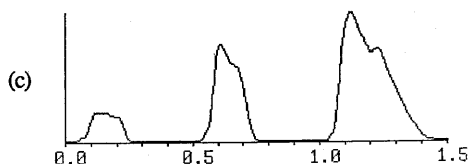
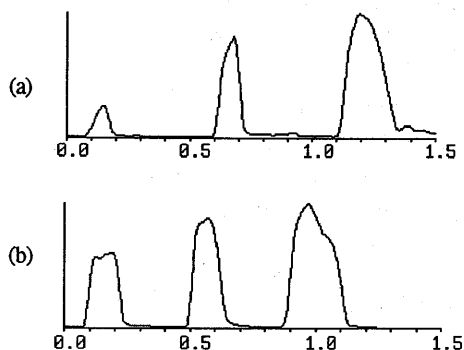


図2 平均振幅の包絡線の比較
(a)鳴き声、(b)鳴き真似1、(c)鳴き真似2)

・基本周波数パターン

平均振幅がしきい値以上となるフレームが80msec連続する部分を有声区間とし、有声区間の8msecごとに32msec区間の基本周波数を自己相関関数により計算した。また無声区間の基本周波数は0とした。図3の(a)はシマアオジの鳴き声の基本周波数パターンであり、(b)と(c)は2人の被験者による鳴き真似の基本周波数パターンである。

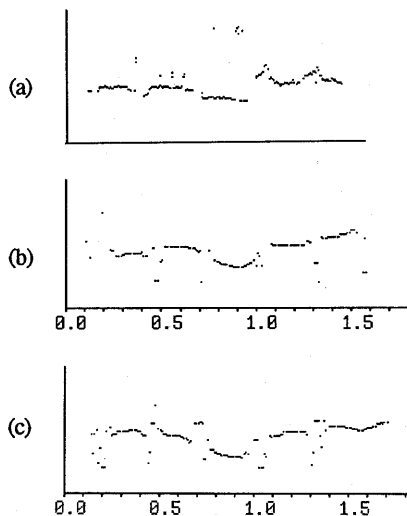


図3 基本周波数パターンの比較
(a)鳴き声、(b)鳴き真似1、(c)鳴き真似2)

3.2 鳴き真似の類似構造

図2を見ると鳴き声と鳴き真似の平均振幅の包絡線は以下の3点で類似している。

- (1) 包絡線の山の数がともに3個ずつである。(有声区間の数)
- (2) (a)と(b)で、1番目と2番目の音の継続時間長は等しく、0.15秒程度であり、最後の音の継続時間は最も長く0.25秒程度である。(各有声区間の継続時間長)
- (3) (a)と(c)で、1番目の音の平均振幅の大きさが

最も小さく最大値の30%程度、次に2番目の音で70%程度、そして最後の音が最大となっている。(各有声区間の平均振幅の大きさ)

図3の鳴き声と鳴き真似の基本周波数パターンにおいても以下の類似点がある。

(1) 有声区間がともに5個存在する。(有声区間の数)

(2) 有声区間の継続時間が全て0.2秒程度で、ほぼ等しい。(各有声区間の継続時間長)

(3) 1番目と2番目の音の基本周波数の大きさは等しく、3番目の音の基本周波数は最も低い。また4番目と5番目の音の基本周波数はほぼ等しく全体で最も高い。(各有声区間の基本周波数の大きさ)

以上まとめると、鳴き声の大きさと高さに関する鳴き真似の類似構造は、(1) 有声区間の数、(2) 対応する有声区間の継続時間長、(3) 対応する有声区間の正規化された平均振幅の値と基本周波数の値、である。

4. 鳴き真似の類似構造に基づく距離の計算方法

本節では、まず鳴き真似の類似構造に基づく鳴き声と鳴き真似の距離を定義する。さらに実際に鳴き真似より検索するアルゴリズムについて述べる。

4.1 距離の定義

鳴き声と鳴き真似の平均振幅の包絡線の距離を両パターンを重ね合わせたときの面積の差と定義すると、第3節で述べた鳴き真似の類似構造は次のように反映される。

(1) 有声区間の数

有声区間の数が鳴き声と鳴き真似で違っていると、無声区間と有声区間を重ね合わせることで、面積の差は増加する。有声区間の数が等しいと、包絡線の山を鳴き声と鳴き真似でそれぞれ対応づけることができ、面積の差は減少する。

(2) 有声区間の継続時間長

対応する有声区間でその継続時間長(幅)を比較する。幅の長さが違うと面積の差は増加し、幅の長さが等しいと面積の差は減少する。

(3) 有声区間の平均振幅の大きさ

対応する有声区間でその平均振幅の大きさ(高さ)を比較する。高さが違うと面積の差は増加し、高さが等しいと面積の差は減少する。

同様に、鳴き声と鳴き真似の基本周波数パターンの距離も両パターンを重ね合わせたときの面積の差と定義する。

4.2 検索アルゴリズム

音の平均振幅の包絡線を振幅構造、基本周波数パターンをピッチ構造と呼ぶ。振幅構造は音の大きさの変化に対応し、ピッチ構造は音の高さの変化に対応する。以下では各構造ごとに、その抽出アルゴリズムと鳴き声と鳴き真似間の比較アルゴリズムについて述べる。

4.2.1 振幅構造

(1) 振幅構造の抽出

鳴き真似の一定区間(フレーム)を取り出し、そのフレームの平均振幅を計算する。フレームを一定周期毎に移動させ、各フレームの平均振幅を求める。さらに全てのフレームの平均振幅が求まるとその最大値で正規化する。

(2) 振幅構造の比較

図2、3の鳴き声と鳴き真似を比較すると、良く類似しているが、時間軸がずれていることがわかる。この時間軸のずれは鳴き声と鳴き真似の本質的な差ではないので、両パターン間の時間軸のずれの影響をできるだけ受けない距離の計算方法が望ましい。時間軸のずれた2つのパターンに対し、時間軸を非線形に伸縮させることにより、2つのパターンの最小距離を求めるDPマッチングのアルゴリズムがある[6]。そこで鳴き声と鳴き真似の距離の計算では、DPマッチングによる比較を応用する。

鳴き声の振幅構造、 $A=A[1], A[2], \dots, A[M]$ 、と鳴き真似の振幅構造、 $B=B[1], B[2], \dots, B[N]$ の距離、 $DISTANCE(A, B)$ は以下のように計算する。

$distance(i, j)$ は、各振幅構造の要素間の距離を表わし、 $abs(A[i], B[j])$ で計算される。 $G[i, j]$ は2次の配列で、 $distance(i, j)$ の累積和に相当する。また極端な時間伸縮を避けるために、整合窓 r を用いる。

(1) $G[i, j]$ の初期化。

$$G[1, 1] = 2distance(1, 1)$$

(2) $j=1$ より N まで(3)を実行。

(3) $i=\max(1, j-r)$ から $i=\min(j+r, M)$ まで $G(i, j)$ を計算。

$$G[i, j] = \min\{G[i, j-1] + distance(i, j),$$

$$G[i-1, j-1] + 2distance(i, j),$$

$$G[i-1, j] + distance(i, j)\}$$

(4) $G[M, N]$ を正規化し、 $DISTANCE(A, B)$ を得る。

$$DISTANCE(A, B) = G[M, N] / (M + N)$$

利用者には、 $DISTANCE(A,B)$ が小さい鳥の順に提示される。

4.2.2 ピッチ構造

(1) ピッチ構造の抽出

まず平均振幅がしきい値以上となるフレームが一定時間連続する部分を有声区間とみなす。その有声区間において、各フレーム毎に自己相関関数により基本周波数(ピッチ)を求める。そしてフレームを一定周期毎に移動させ、各フレームのピッチを求める。また無声区間では、ピッチを0とする。

(2) ピッチ構造の前処理(正規化)

DPマッチングによる比較の方法が行なえるように、基本周波数パターンを折れ線で近似する。

(a)誤差の処理

自己相関関数によるピッチの計算では、正しいピッチではなく、その倍数(倍ピッチ)や半分のピッチ(半ピッチ)を取り出す抽出誤差がみられる。ところが人の音韻では極端にピッチが変化することはない、同様に鳥の鳴き声においても1つの有声区間で極端にピッチが変化することはない。そこで1つの有声区間で他のピッチとの距離の平均が最も短いピッチ(中央値)を求め、中央値との差が中央値の2分の1以上となるピッチを誤差とする。そして誤差となる部分を前後の誤差とならないピッチで線形補間する。

(b)折れ線近似

縦軸がピッチ、横軸が時間であるグラフを考え、まず有声区間の始点と終点を直線で結ぶ。各時間の近似値をその時間における直線のピッチの値とする。近似値と実際の値の差の平均を求め、しきい値以上の場合には最初に差がしきい値を越える点を始点として、同じ処理を繰り返す。近似値と実際の値の差の平均がしきい値を越えない場合はこの有声区間の折れ線近似はできたものとし、次の有声区間を処理する。

全ての有声区間の折れ線近似が終了すると、全区間のピッチをその最大値で正規化する。

(3) ピッチ構造の比較

折れ線で近似されたピッチ構造は、振幅構造と同様に線形なパターンとなるので、ここでも先に説明した振幅構造の比較アルゴリズムと同様の方法で比較する。

5. 評価実験と考察

20種類の鳴き声に対する各52人の鳴き真似、合計1040個の鳴き真似を用いて、4節で述べた検索アルゴリズムの評価実験を行った。評価実験の目的は、鳴き真似の類似構造を検索アルゴリズムが反映しているか調べることである。本節では、評価実験の概要とその結果についてまず説明し、さらにその結果を考察する。

5.1 実験概要

(1) 実験データの収集

・実験に用いた鳴き声データ

野鳥50種類の鳴き声を収録したCD[7]より、「同じ鳴き声を繰り返す鳥を集める」点に注意し、20個の評価データを定めた。表1に評価データの一欄を示す。表中の時間は1つの鳴き声の継続時間(秒)を現す。また備考は、各鳴き声に対するそれぞれ52個の鳴き真似を著者の一人が聞き、「この鳥ではなからうか」と判断できた割合を3段階評価したものである。「BEST」が70%以上判定できた場合、「GOOD」が70%から30%の割合で判定できた場合、「BAD」が30%以下の割合でしか判定できなかった場合を意味する。

表1 実験に用いた鳴き声データ

鳴き声	時間	備考
カッコウ	0.73	BEST
トラツグミ	0.79	BEST
イソシギ	0.89	BEST
アオアシシギ	0.93	BEST
コノハズク	1.51	BEST
ホトトギス	1.02	GOOD
アカハラ	1.22	GOOD
コマドリ	1.36	GOOD
ウグイス	1.50	GOOD
アカショウビン	1.53	GOOD
シマアオジ	1.64	GOOD
シジュウカラ	1.92	GOOD
ヒガラ	2.06	GOOD
ヤマガラ	2.16	GOOD
コガラ	2.59	GOOD
メボソムシクイ	2.94	GOOD
ヤブサメ	4.01	GOOD
ルリビタキ	1.33	BAD
イカル	1.49	BAD
ホオジロ	1.54	BAD

- ・被験者
52人の男女学生（男子34人、女子18人）
- ・実験データの収集
被験者に1つの鳴き声につき5回程度聞かせ、その鳴き真似をサンプリング周波数8KHzで録音した。
- ・実験データの前処理
収集した1040個の鳴き真似に対し、実際に鳴き真似を聞きながら、またその波形を見ながら雑音部分を除去した。
- ・実験のパラメータ
フレーム長は32msec、フレーム周期は8msecとした。

（2）評価実験

[実験1]

実験1では、振幅構造で比較する検索を評価した。また第3節の鳴き声と鳴き真似の音響分析による比較を見ると、鳴き声と鳴き真似ばかりでなく鳴き真似同士も類似していることがわかる。そこでシステムデータとして、（1）実際に被験者に聞かせた鳥の鳴き声の振幅構造を用いた場合、（2）任意に選んだ7人の被験者内で他と最も距離が短い鳴き真似の振幅構造を用いた場合、で評価した。

[実験2]

実験2では、ピッチ構造で比較する検索を評価した。鳴き声によってピッチ構造の抽出が難しい鳥があるので、任意に選んだ7人の被験者内で他と最も距離が短い鳴き真似のピッチ構造をシステムデータとした。

5.2 実験結果

2節で述べたように、検索結果として利用者には、入力した鳴き真似との距離が短い鳥の順に提示される。そこで評価では、被験者が鳴き真似の対象とした鳥を最も距離の短いものとして検索できた場合ばかりでなく、距離が短い順に5位以内に鳴き真似の対象となった鳥が入る場合も表示した。

[実験1]

図4は実験1（1）の結果を示す。縦軸は鳥の種類を表し、横軸は正しい検索結果を1位に提示される利用者の割合と正しい検索結果を5位以内に提示される利用者の割合を表す（検索成功率と呼ぶ）。これによると、平均して31%の被験者が正しい検索結果を1位に提示され、67%の被験者が5位以内で提示される。また図5は実験1（2）の結果である。これによると正しい検索結果を1位に提示される利

用者の割合は46%で、正しい検索結果を5位以内に提示される利用者の割合は77%となっている。

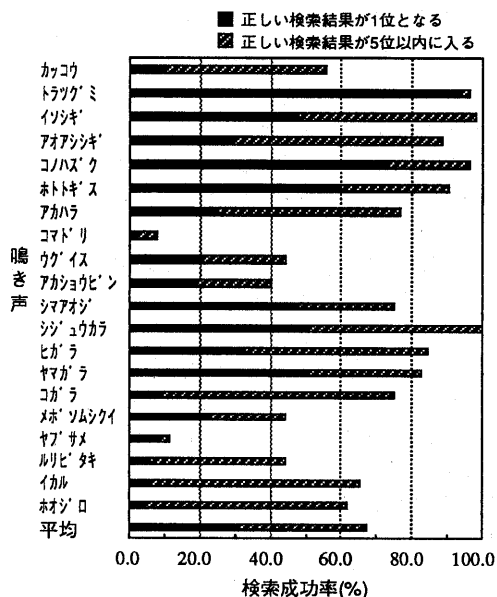


図4 振幅構造で比較する検索の評価
(鳴き声をシステムデータとした場合)

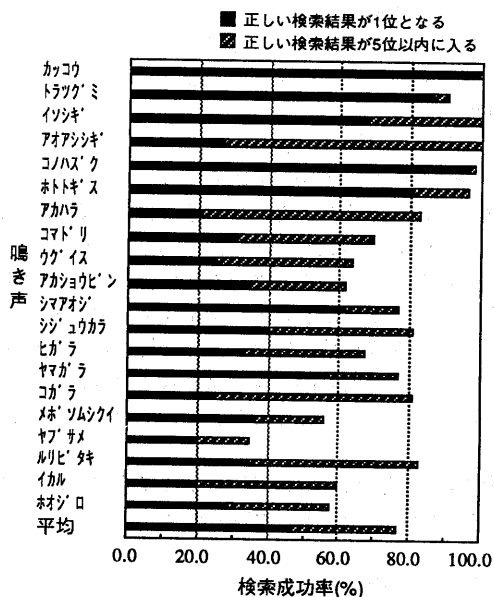


図5 振幅構造で比較する検索の評価
(鳴き真似をシステムデータとした場合)

[実験2]

図6には実験2の結果を示す。これによるとピッチ構造で比較する検索では、37%の被験者が正しい検索結果を1位に提示され、82%の被験者が5位以内で提示されることがわかる。

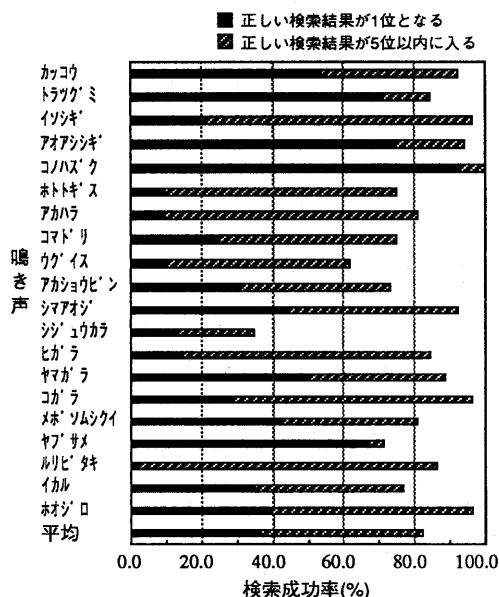


図6 ピッチ構造で比較する検索の評価
(鳴き真似をシステムデータとした場合)

5.3 考察

(1) 鳴き真似の類似構造の反映について

図7(a)はアオアの鳴き声の振幅構造であり、図7(b)、(c)は2人の利用者のアオアの鳴き真似の振幅構造である。(a)と(b)の距離は2.71で、他の鳴き声と(b)との距離と比べて最も短い。一方、(a)と(c)では有声区間の数が違うので、距離は大きくなる。(a)と(c)の距離は5.70で、他の鳴き声と(c)との距離と比べて8番目に短い。

同様に、他の類似構造が反映されている鳴き真似と反映されていない鳴き真似を比較すると、前者の鳴き真似の方が元となった鳥との距離が短い。提案した検索アルゴリズムで、利用者が鳴き真似の対象とした鳥を最も距離の短いものとして検索できる場合、その鳴き真似は(1)鳴き声を構成する音の数、(2)鳴き声のリズム、(3)鳴き声の大きさと高さを表現している。

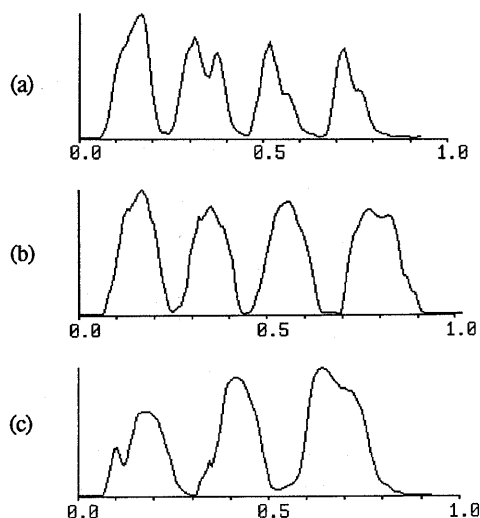


図7 アオアの振幅構造の比較
(a)鳴き声、(b)鳴き真似1、(c)鳴き真似2)

(2) 人の直感と検索アルゴリズムの関係について
第4節で提案した検索アルゴリズムが人の直感的な類似判断基準に基づいているかについて考察する。表1に示した人の直感による検索率が、「BEST」「GOOD」「BAD」となる鳥ごとに実験1(1)の検索成功率の平均を求めた。人による検索率は、「BAD」「GOOD」「BEST」の順に上がるが、同様に検索アルゴリズムによる検索成功率も上昇している。

表2 人による検索率と検索アルゴリズムによる検索成功率の関係

(表中の数字は検索成功率を表す(%))

人	検索アルゴリズム	
	1位	5位以内
BEST	51	87
GOOD	34	61
BAD	5	57

また実験に用いた鳴き声のうち、ヒガラとシジュウカラ、アオアとイソシギは直感的に類似している。第2節で述べたように、利用者には入力した鳴き真似に近い鳥の順に提示される。人の直感による類似判断基準を反映できる検索システムであるならば、例えばヒガラの鳴き真似を検索要求とした場合でもシジュウカラは検索要求と類似しているとして利用者に提示されるはずである。そこで実験1(1)の場合において、それぞれ直感的に類似した鳴き声が上位5位以内に提示される割合を求め、これを表3

に示す。それぞれ高い割合で、人が実際に似ていると感じる鳴き声をもつ鳥が5位以内で利用者に提示される。

表3 人が似ていると感じる鳴き声が上位5位以内に入る割合

鳴き真似	類似した鳴き声	割合 (%)
ヒガラ	シジュウカラ	94
シジュウカラ	ヒガラ	83
アオア	イソシギ	75
イソシギ	アオア	85

以上表2、表3の結果より、提案した検索アルゴリズムには、人の直感的な類似判断基準が反映されていることがわかる。

(3) 検索が失敗する理由と今後の課題について

評価実験で検索成功率が低い鳥についてその理由を考察する。以下は、特に実験1(1)で検索成功率が低い鳥とその特徴である。尚、この検索が失敗する場合についての対策は、今後の課題である。

・継続時間が長い鳴き声

図3より、ヤブサメとメボソムシクイの検索成功率はともに低いことがわかる。表1をみると、ともに鳴き声の継続時間が3秒から4秒と他と比べて長い。またヤブサメとメボソムシクイの鳴き真似の継続時間を調べると、鳴き真似ごとにまちまちである。この事実より、継続時間が長い鳴き声に対して被験者は、正確にその継続時間を鳴き真似で表現できないと考えられる。今回提案した検索アルゴリズムでは、DPマッチングでの極端な時間軸の伸縮を防ぐため、しきい値以上で継続時間が違うと距離が最大となる。今後、システムデータと鳴き真似の継続時間が極端に違う場合の比較の方法が必要となる。

・音の大きさや高さの変化が速い鳴き声

図8にホオジロの鳴き声と鳴き真似の振幅構造を示す。ホオジロは継続時間が1.5秒程度で、その短い時間の間に激しく変化している。一方、被験者の鳴き真似は中途半端で鳴き声の激しい変化を真似ることができていない。このことより、被験者は音の大きさが激しく変化する鳴き声を鳴き真似で正しく表現できないと考えられる。コマドリとルリビタキも同様の理由で検索成功率が低い。考察(1)で述べたように、提案した検索アルゴリズムでは、鳴き真似は鳴き声を構成する音の数、鳴き声のリズム、

鳴き声の大きさと高さを表現していないと検索できない。そこで今後、他の特徴、例えば音色の変化など、に着目した検索アルゴリズムが必要となる。

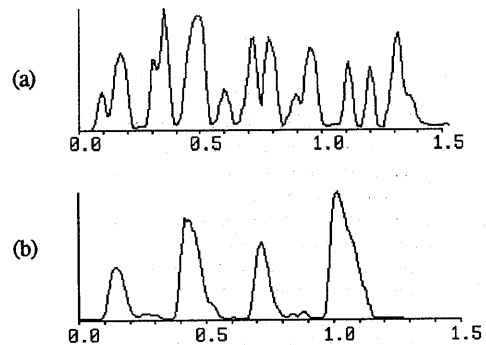


図8 ホオジロの鳴き声と鳴き真似の振幅構造
(a)鳴き声、(b)鳴き真似)

・個人の先入観がある鳴き声、人と発声の仕組みが違う鳴き声

一般にカッコウやウグイスは、良く知られた鳥で鳴き真似も容易であると考えられる。ところが図3を見るとカッコウの検索成功率は5位以内で55.8%、ウグイスの検索成功率は44.2%であり他と比べて低い。図9にカッコウの鳴き声と鳴き真似の振幅構造を示す。これによると両者の違いは第2音に見られる。実際のカッコウの鳴き声では第1音は音が大きく、第2音はほとんど聞こえない。しかし被験者にはカッコウの鳴き声は2つの音ともにはっきり大きく聞こえるという先入観があり、正確に鳴き真似できない。

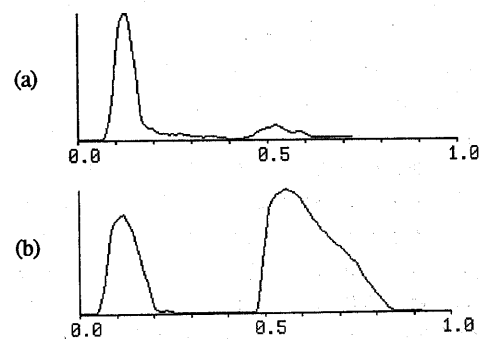


図9 カッコウの鳴き声と鳴き真似の振幅構造
(a)鳴き声、(b)鳴き真似)

カッコウの場合は、手本となる鳴き真似をシステムデータとすることによりこの課題を解決できた(図4、5)。ところがウグイスの場合は、図4、5を見てもうまく解決できていない。図10はウグイスの鳴き声と鳴き真似の振幅構造である。ウグイスは「ホーホケキョ」の聞きなしで知られているが、「ホー」の音に対応する第1音は、吸気音であり人は吸気音を出すことはできない[8]。またこの吸気音は人により聞こえる場合と聞こえない場合があるようで、利用者ごとに第1音を省いて鳴き真似したり、第1音を強調して鳴き真似したりしてる。そこで今後、人と鳥との発声の違いに着目した検索方法が必要となる。

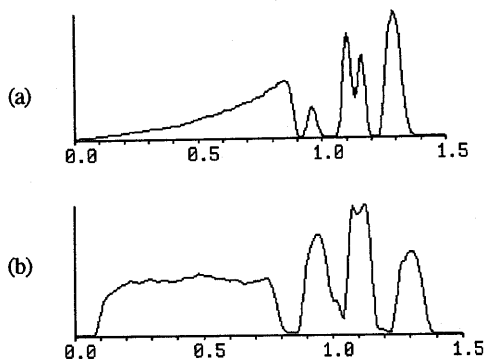


図10 ウグイスの鳴き声と鳴き真似の振幅構造
(a)鳴き声、(b)鳴き真似)

6. まとめ

本稿では、鳥類図鑑Hyperbookにおける鳴き真似による検索を実現するために、鳴き真似の類似構造とそれに基づく検索アルゴリズムについて述べた。本稿で解明した鳴き真似の類似構造は、(1)鳴き声を構成する音の数、(2)鳴き声のリズム、(3)鳴き声の大きさと高さである。また検索アルゴリズムではこれらの類似構造を反映するために、音の大きさと高さのパターンを連続したパターンと考え、それぞれの鳴き声と鳴き真似の距離をパターンのなす面積の差と定義した。また比較の方法ではDPマッチングを応用した。

20種類の鳥の鳴き声に対するそれぞれ52人の鳴き真似、計1040個の鳴き真似を用いた評価では、(1)振幅構造で比較する検索で、検索候補を距離の短い順に並べたとき、正しい検索結果が1位となる割合が平均46%、上位5位以内に入る割合が平均77%、(2)ピッチ構造で比較する検索で、1位となる割合が平均37%、上位5位以内に入る割合が平均82%

であった。実験に用いた鳴き真似データの中で我々が聞いてもどの鳴き声の真似をしているのかわからない場合があることを考えると、今回提案した検索アルゴリズムが、本稿で示した類似構造を利用者が表現しやすい鳴き声の検索において有効であるといえる。

今後の課題であるが、検索ができなかった鳴き声と鳴き真似の例について考察を進め、人が鳴き声の何を真似ているかをさらに検討する。今回の鳴き真似の類似構造の解明では、音の音色は物理的にどの特性と対応しているかわからないため省いて考えた。現在まで、例えば鳴き真似を口真似で行なった場合、その鳴き真似の音韻の変化が鳴き声のスペクトルの変化に対応していることがわかっている。そこで今後、鳴き声のスペクトル変化構造に着目した類似構造を明らかにする。

謝辞

本研究を進めるにあたり、日頃より御討論下さる早稲田大学理工学部村岡研究室の上野正俊君と藤沢正幸君に感謝します。また実験に御協力くださった早稲田大学理工学部、日本女子大学、日本児童教育専門学校の学生の皆様に感謝します。

参考文献

- [1]M.Tabuchi,Y.Muraoka:"MeSOD:the Metric Spacial Object Data model for a multimedia application:Hyperbook", IEEE Compcon89 spring,1989.
- [2]中村,行田:"野鳥検索小図鑑",講談社,1986
- [3]藤沢,矢川,田淵,村岡:"鳥類図鑑Hyperbookシステムにおける対話型検索について",情報研資料89-DB-74-5.
- [4]矢川,上野,田淵,村岡:"鳥類図鑑Hyperbookにおける鳴き声検索",情報40回全大4H-3,1990.
- [5]"聴覚ハンドブック",ナカニシヤ出版社,pp171-180.
- [6]新美:"音声認識",共立出版,1979
- [7]"bird song",主婦の友社,1990
- [8]川村:"鳥の歌の科学",中央公論社,1974