

# 鳥類図鑑Hyperbookにおける 7M-2 鳴き真似を用いた検索の実現方式について

矢川 雄一 田淵 仁浩 村岡 洋一  
早稲田大学 理工学部

## 1. はじめに

距離空間データモデルMeSOD[1]は、実世界の実体間の関連度を定義することにより、実世界における曖昧な検索を可能にするデータモデルである。現在、我々はこのMeSODに基づき、利用者の自由な問い合わせを可能とする鳥類図鑑Hyperbook[1]を研究、開発中である。

本稿では、鳥類図鑑Hyperbookにおける利用者の自由な問い合わせとして鳴き真似を用いた検索を取り上げ、その実現方法について述べる。

## 2. 鳴き真似を用いた検索の実現可能性

ここでは、鳴き真似を用いた検索について述べたあと、鳴き声と鳴き真似の類似点を明らかにし、鳴き真似を用いた検索が実現可能であることを説明する。

### 2.1 鳥類図鑑Hyperbookの鳴き真似を用いた検索

鳥の検索に有効な情報として鳴き声があるが、鳥類図鑑Hyperbookにおいては、この鳴き声情報を利用した検索として、鳴き真似を用いた検索を考えている。[2]この鳴き真似を用いた検索は、鳥類図鑑Hyperbookにおけるマルチメディアの利用例であり、音声メディアを検索要求の記述に許す技術として意義がある。

### 2.2 鳴き真似の音韻情報と韻律情報

利用者の鳴き真似を検索要求とするとき、鳴き真似は鳥の鳴き声と決して一致しないが、いくつかの類似点が見られる。すなわち以下に述べるような音韻情報(直感的には鳴き声の音色に対応する)と韻律情報(抑揚、メロディ、リズム)で、鳴き真似は鳴き声と類似している。さらに同じ鳥の鳴き声に対する利用者毎の鳴き真似も音韻情報、韻律情報で類似している。

#### (1)音韻情報

表1はツアツの鳴き真似の音韻情報の例である。これを見ると、それぞれ用いている音素には差違があるが(音の置換)、音素のパターンは類似している。すなわち1番目と2番目の音素、4番目と5番目の音素は同じであるのに、3番のみは他と違っている。また実際のツアツの鳴き声のスペクトル構造(音色構造)をみると、スペクトルの変化の様子が鳴き真似の音素のパターンと類似している。

表1 鳴き真似における音韻情報

被験者	鳴き真似	被験者	鳴き真似
A	ト <sup>°</sup> イ <sup>°</sup> -ヒ <sup>°</sup> -イ <sup>°</sup> -ヒ <sup>°</sup> ヨ <sup>°</sup> ヒ <sup>°</sup> イ <sup>°</sup> イ	C	チ <sup>°</sup> イ <sup>°</sup> -チ <sup>°</sup> -ヒ <sup>°</sup> ヨ <sup>°</sup> チ <sup>°</sup> イ <sup>°</sup> イ
B	ヒ <sup>°</sup> -ヒ <sup>°</sup> -ヒ <sup>°</sup> ヒ <sup>°</sup> ヨ <sup>°</sup> ヒ <sup>°</sup> イ <sup>°</sup> イ	D	ヒ <sup>°</sup> イ <sup>°</sup> -ヒ <sup>°</sup> イ <sup>°</sup> -ホ <sup>°</sup> ヒ <sup>°</sup> イ <sup>°</sup> イ

#### (2)韻律情報

鳴き声と鳴き真似の韻律情報を比較すると、鳴き声で観察される音が鳴き真似では観察されないという問題(音の脱落)も見られたが、両者は類似していた。また同じ鳥の鳴き声に対する利用者毎の鳴き真似の韻律情報はさらに類似していた。(図1)尚、ここで韻律情報とは振幅構造(抑揚)、ピッチ構造(メロディ)、時間構造(リズム)である[3]。

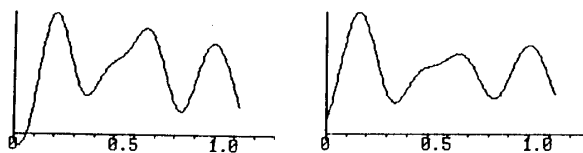


図1 ツアツの鳴き真似の振幅構造の比較

### 2.3 実現の可能性

上述したように鳴き真似の音韻情報と韻律情報に鳴き声との類似点が見られる。また同じ鳥の鳴き声に対する利用者毎の鳴き真似も類似している。この2点を利用して鳴き声と鳴き真似の関連度を定義すれば、鳴き真似を用いた検索をMeSODの枠組みで捉えられる。

## 3. 鳴き真似を用いた検索の実現方式

ここでは上述した類似点を利用して、鳴き真似を用いた検索を実現する方法について述べる。

### 3.1 実現のための課題

検索を実現するための方法として、従来の音声認識の手法を応用する方法があるが、次に述べる問題を考慮しなくてはならない。

#### (1)音韻情報ばかりでなく韻律情報も積極的に扱う

2.2で述べたように、鳴き真似では音韻情報ばかりでなく、韻律情報も検索を行なう上で有効な情報である。ところが従来の単語音声認識では、韻律情報を積極的に認識のために用いる工夫はなされていない。

#### (2)音の脱落、置換による影響を吸収する

鳴き真似を観察すると、利用者によって鳴き声ではみられた音の脱落があったり、利用者毎に鳴き真似に用いる音が違う(音の置換)といった問題もみられる。

### 3.2 課題解決の方針

単語音声認識で、音の置換、脱落を扱う方法として単語を確率モデルで表現する方法がある。[4][5]この確率モデルでは、音の置換や脱落を確率の問題として扱えるので上述した(2)の問題点に有効である。

また従来の確率モデルを用いた単語音声認識では、モデルの訓練、評価を行なうのに音韻情報を用いているが、(1)の問題に対処するために、確率モデルの訓練、評価には音韻情報、韻律情報ともに用いることとする。

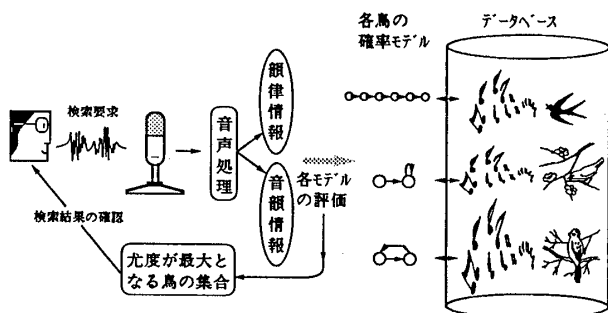


図2 鳴き真似を用いた検索の実現方式

### 3.3 具体的な実現方法

ここで具体的な実現方法について述べる。(図3)

#### (1)鳴き声グラフの構成

鳥の鳴き声ごとに図4のような鳴き声グラフを構成する。これは鳴き声の音色パターン(スペクトル構造)、振幅構造、ピッチ構造、時間構造の変化の様子を考慮して構成され、鳴き声グラフの各枝は、鳴き声の各構造が安定している部分に対応する。例えば図4の①ツアツアでは安定した部分は5つであり、それぞれグラフの各枝に対応している。尚、アカショウビンは2番目の音以降、同じ構造が連続するのでこのようなグラフとなる。

#### (2)確率モデルの訓練

鳴き声グラフが求まると、次に複数の鳴き真似を用いて訓練を行ない、確率モデルを得る。訓練用の鳴き真似はまず音声処理が施され、音韻情報(音素パターン)を表す音韻系列と韻律情報(振幅構造、ピッチ構造、時間構造)を表す韻律系列が得られる。そして鳴き声のスペクトル構造、振幅構造、ピッチ構造、時間構造を参照しながら、鳴き真似の音韻系列と韻律系列による訓練が行なわれる。この訓練を複数の鳴き真似を用いて行ない、鳴き声の確率モデルの状態遷移確率(音の脱落の影響を吸収する)、音韻情報の出力確率(音の置換の影響を吸収する)、韻律情報の出力確率を得る。

#### (3)関連度(尤度)の計算(図5)

まず利用者の鳴き真似から音韻系列、韻律系列を得る。次に候補となる鳴き声の確率モデルでこの韻律系列を出力する経路を全て求め、その出力確率を求める。最大の確率を与える経路を最適経路とし、今度はこの経路で検索要求の音韻系列を出力する確率を求める。ここで韻律系列による最適経路の確率と音韻系列の出力確率が求まるが、候補の確率モデルにおける利用者の鳴き真似の尤度は、両

確率を掛け合わせたものとし、この尤度を鳴き声と鳴き真似の関連度とする。

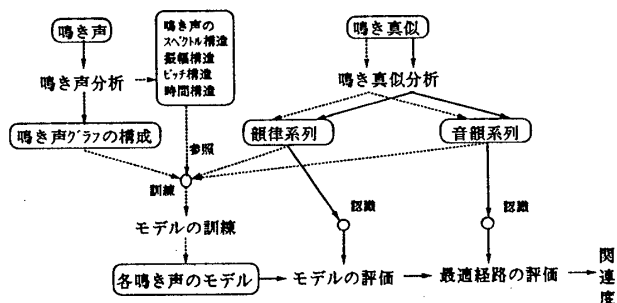


図3: 確率モデルによる具体的な実現方法

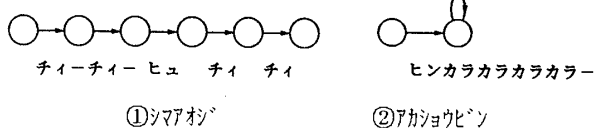


図4: 鳴き声グラフの例(下は代表的な鳴き真似例)

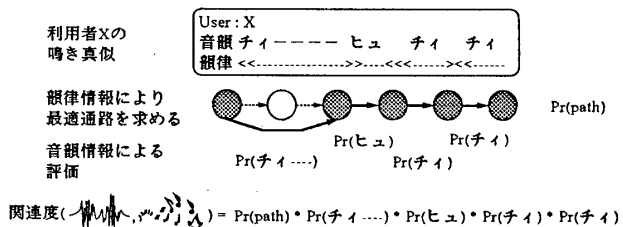


図5: 尤度の計算

鳴き真似と鳴き声との関連度(尤度)を以上の操作で求めることにより、利用者は関連度が大きいほうから順に3つの鳥を取り出し、実際に鳴き声を聞くといった検索結果の確認も可能となる。

### 3.3 実現方式の評価

実際の鳥の鳴き声と20人の学生による鳴き真似をモデルの訓練に用い、別の1人の鳴き真似を用いて評価をおこなっており、現在まで良好な結果を得ている。

### 4. おわりに

本稿では、鳥類図鑑Hyperbookにおける利用者の自由なメディア利用として、鳴き真似を用いた検索を取り上げ、その実現方式について述べた。確率モデルを用いた関連度を定義することにより検索対象と決して一致することがない鳴き真似を用いて検索が可能である。

#### [参考文献]

[1]M.Tabuchi,Y.Muraoka:"MeSOD:the Metric Spacial Object Data model for a multimedia application:Hyperbook",IEEE Comcon89 spring,1989.  
 [2]芝崎,柴山,田淵,村岡:"距離空間データモデルMeSODにおけるメディア検索について",情処36回全大2E-3,1988.  
 [3]北原,東倉:"音声の情緒表現に關与する韻律成分",音響講義1-2-15,1988.  
 [4]中川:"確率モデルによる音声認識",電子情報通信学会,1988  
 [5]F.Jelinek:"Continuous speech recognition by statistical methods",Proc.IEEE,64,1976.