

E-075

K-means法を用いたジュウシマツの歌の音素分類に関する研究

Phonetic analysis of Bengalese finch's song using K-means clustering

小松 靖直[†]
Yasunao Komatsu

柿下 容弓[‡]
Yasuki Kakishita

笹原 和俊[§]
Kazutoshi Sasahara

西野 哲朗[¶]
Tetsuro Nishino

高橋 美樹[§]
Miki Takahasi

岡ノ谷 一夫[§]
Kazuo Okanoya

1 はじめに

歌鳥と呼ばれる種の鳥は後天的に音声学習する能力を持つことが知られている。これらの鳥の歌の学習過程と人間の音声言語獲得過程には、多くの共通点があることから、歌鳥は人間の音声言語獲得過程の良いモデル生物として扱われている。中でもジュウシマツは特に複雑な歌を学習する種として注目され、盛んに研究が行われている [1]。

歌を解析する多くの場合、音素分類を行って得られるテキストデータを使用する。しかしジュウシマツの歌の音素分類を機械的に行う手法は確立しておらず、人の手で行われているのが現状である。そこで我々は、Tchernichovskiらの先行研究を参考に WienerEntropy と音圧を用いた音素境界の推定を行い [2]、K-means法によるクラスタリングを音素に応用することで音素分類作業の自動化、効率化を目指した。

2 ジュウシマツの歌

ジュウシマツの歌は1バウト(1うたい)あたり、100個前後の音素で構成される。図1はジュウシマツの歌の音声波形とそのソナグラムである。ソナグラムとは対象音声のある時刻における周波数成分の強さ $S(t, f)$ を色の濃淡で表したものである。

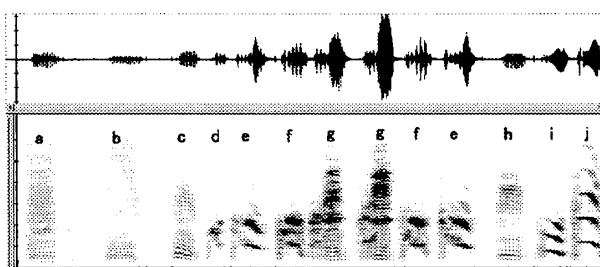


図1: (上)ジュウシマツの音声波形, (下)ソナグラム 横軸:時刻 [s](目盛幅 0.05s), 縦軸:周波数 [Hz](目盛幅 2kHz)

ジュウシマツの場合、1つ1つの音素は無音区間で区切られる。そのためソナグラム上での一つの塊を音素とみなすことができる。同一個体において、同一形状と考えられる音素には同

[†]東京工業大学大学院 総合理工学研究科
[‡]電気通信大学大学院 電気通信学研究科
[§]理化学研究所 BSI 生物言語研究チーム
[¶]電気通信大学 情報通信工学科

じ記号を割り当てることで歌を記号列として表現することができ、この作業を音素分類と呼ぶ。例えば図1の表示内では、音素数が13個、得られる記号列は「abcdefghgfehij」である。

3 提案手法

本論文ではジュウシマツの歌の音素分類を、音素境界の推定と音素分類の二段階に分けて考える。本節ではこのそれぞれについて詳しく述べる。

3.1 音素境界推定

音素分類を行うためには、まず各音素の始点と終点を推定し、音素区間を得る必要がある。本研究では2つの音声特徴量を用いて、無音区間か音素区間か判別することでこれを行った。

Amplitude: ある時刻における信号の強さを表す量

WienerEntropy: 信号のランダムさの程度を表す量。音声雑音に近いほど1に近く、規則的なほど0に近くなる

両音声特徴量には音素区間で値が大きくなる傾向がある。そのためある閾値を定め、それを越えた1区間を1音素とみなすことで音素境界を推定した。図1にAmplitudeを適用した例を図2に示す。縦線が音素の境界を表している。

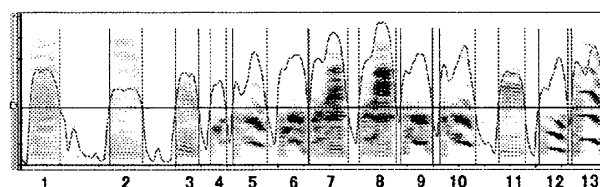


図2: Amplitude適用例

本研究では、汎用性を高めるために2つの音声特徴量を組み合わせることとした。すなわち以下の4つの内、最も適したものを選択する。

1. Amplitude
2. WienerEntropy
3. Amplitude AND WienerEntropy
4. Amplitude OR WienerEntropy

ただし単純に閾値を越えた区間を音素区間とするだけでは不十分である。何故なら音声データの中にはクチバシがぶつか

る音等、ノイズ音も含まれているからである。そこでこのような状況に対処するためある一定の幅以下の音素を歌以外の音と見なし、分割の対象から外すこととした。

3.2 K-means 法による音素分類

得られた音素を分類するために、各音素の類似度を求めなければならない。そこで本研究では、それぞれの音素区間から幾つかの特徴量を導出し、K-means 法によりクラスター解析を行うこととした。クラスター解析には統計解析システム R を用いた。

音素分類を行うための特徴量として、3.1 節で述べた Amplitude と WienerEntropy に加え、以下の 4 つを使用した。これらの定量化法を適用し、それぞれの音素区間における平均値を特徴量とした。ただし音素の中には途中で周波数成分が変化するものもあるため、音素の前半と後半それぞれの平均値を求め、特徴量とすることで精度向上を目指した。

AmplitudeModulation : 音声信号の振幅の変化量

FrequencyModulation : 周波数成分の変化の度合いを表す量

HarmonicPitch : ソナグラム上での主要な周波数

Mean Frequency : $S(t, f)$ のある時刻 t における周波数成分の平均値

これらに各音素の時間幅を加えた 13 個の特徴量のセットを 1 つの入力 (特徴量ベクトル) としてクラスター解析を行った。

4 結果と考察

人が音素分類した結果と提案手法で音素分類した結果を比較して一致率を調べた。ここでは、ジュウシマツの歌の 1 バウト分に対する結果の一例を紹介する。尚、人による音素分類は 2 人が別々に音素分類作業を行った後、議論を行った上で判断されたものである。1 バウトに含まれる音素の個数は 96、音素の種類は 10 であった。

まず 3.1 節の手法で音素境界推定を行った。ここで紹介する例については 4. Amplitude OR WienerEntropy”を選択することで全ての音素区間を正しく得ることに成功した。ただし、音素同士の間隔が極端に狭い歌の場合、うまく推定できないものもあった。これは 2 つの音素間に十分な間隔が無いこと 1 つの音素とみなしてしまうことが原因であった。

次に 3.2 節の手法で音素分類を行った。K-means 法は乱数の初期値により結果が変わるため、初期値を変えながら同じ特徴量ベクトルを使用して 100 回クラスター解析を行い、各 run 毎に人が行った音素分類の結果と比較した。その結果、最高 92 %、平均 75 % の割合で人の手によるものと一致することが分かった。

以下に音素分類の結果の一例を示す。上が人による音素分類、下が提案手法による音素分類の一部である。尚、不一致の音素には下線が引いてある。

```
acddeffghefghijklfddffghighijklfdefghighijklfdefg
jkckkehghghijklfkkehghighijklfkehghighijklfdefg
```

この例を見て分かる通り、一致しない音素の種類は限られている。k と d、f と h 等がその代表的な例であり、これらの音素はほとんどの解析で不一致となる傾向にあった。この 4 つの音素のソナグラムを図 3、図 4 に示す。

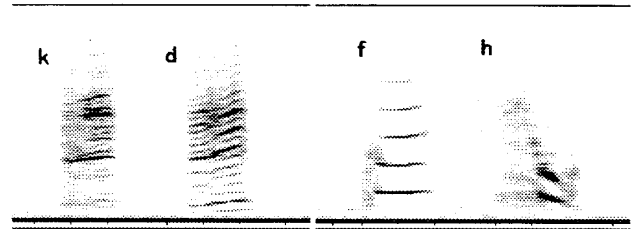


図 3: 音素 k と d

図 4: 音素 f と h

図 3 を見ると k と d は非常に良く似た音素であることがわかる。人が音素分類を行う場合、その音素がどのような文脈で使われるか等も考慮されるため区別されているが、同一の音素と判別しても不自然では無いと考えられる。

一方、f と h に関しては明らかに異なる音素である。導出された特徴量を比較してみると、確かに大きな違いは無かった。今回用いた特徴量だけでは区別しきれない音素の例である。

5 おわりに

本論文では、ジュウシマツの歌の音素境界を推定し、K-means 法を用いてクラスター解析することで自動的に音素分類を行う手法について提案、検証した。

音素境界推定については、1 バウト分のデータに対しても有効であることが分かった。しかし音素同士の間隔が極端に狭い歌をうたう個体に対しては、課題が残る。

次に音素分類に関して述べる。本研究では K-means 法を採用し、人の手による結果と平均で 8 割程度一致するという結果を得た。しかしながら、この手法は初期値の選択次第で適切な結果が得られないケースもある。そのため複数回の解析を行う等、最終的な判断を決める方法を検討する必要がある。また K-means 法は予めクラスタの数を指定しなければならない。効果的な音声特徴量を選ぶことが出来たら、より実用に適したクラスタリング手法を検討する必要がある。クラスター解析に用いる特徴量に関しても、明らかに異なる音素を同一と判断してしまうケースに対応するために、さらなる検討が必要である。

参考文献

- [1] 岡ノ谷 一夫, “小鳥の歌からヒトの言葉へ”, 岩波書店, 2003.
- [2] Tchernichovski, O., Nottebohm, F., Ho, C.E., Bijan, P., Mitra, P.P. (2000). A procedure for an automated measurement of song similarity. *Animal Behaviour* 59, 1167-1176.
- [3] 宮本 定明, “クラスター分析入門 ファジィクラスタリングの理論と応用”, 森北出版株式会社, 1999.