

## 1. 緒言

今日ではスマートフォンでの音声認識が積極的に活用されている。そこでスマートフォンで身近な生き物である鳥の鳴き声で音声認識を行うことを考えた。鳥の鳴き声に音声認識を行うことで、鳥の種類を特定することが目的である。

## 2. 研究のアプローチ

音声認識は「元データ」→「前処理部」→「特徴抽出部」→「認識部」→「識別結果部」の流れとなっている<sup>[1]</sup>。次から具体的な内容を示す。

元データ部では鳥の鳴き声を録音することでアナログ波からデジタル波にする。

前処理部では音声データにフーリエ変換をかけやすくするためにテキストファイルに変換する。

特徴抽出部では、テキストファイルに変換した音声データを配列に入れフーリエ変換を行う。また録音した音声データの音量の入った配列を正規化するために音の最大値で配列を割り、その値を10倍することで比較する値の有効桁数を増やしている。行うフーリエ変換は離散フーリエ変換となっている。

離散フーリエ変換を行うことで音声のスペクトルを実部と虚部に分けて得ることができる。実部に  $\cos$  波のパワーが実数平面上に、虚部には  $\sin$  波のパワーが虚数平面上に表される。この式は式(1)のようになる。

$$F(\omega) = \sum_{n=0}^{N-1} f_n \cos \frac{2\pi kn}{N} - j \sum_{n=0}^{N-1} f_n \sin \frac{2\pi kn}{N} \quad \dots (1)$$

音の特徴は実部と虚部の値の2乗値で表すことができる。これをパワースペクトルと呼ぶ。パワースペクトルを求める式は式(2)のようになる。

$$C_k = ak^2 + bk^2 \quad \dots (2)$$

ここで、 $C_k$  はパワースペクトルの値、 $ak$  はスペクトル実部、 $bk$  はスペクトル虚部である。

認識部では、フーリエ変換を行い求めたパワースペクトル  $C_k$  を配列 A に入れ、基準とした音データを配列 B に入れる。ここで判断式を用いる。配列 B の値と配列 B に一定値  $x$  を足した値の範囲内に配列 A の値となれば判断式が成立する。判断式は式(3)となる。

$$B \leq A \leq B + x \quad \dots (3)$$

この判断式(3)が成り立つならば一致数を1つカウントする。一致数が一定値を超えると基準とした音声データの鳥となる。

識別結果では一致数が一定値を超え基準とした音データの鳥の種類を表示する。

以上の流れで鳥の種類を特定が完了する。

## 3. 結果

今回の実験では、身近な鳥であるスズメを特定することにした。スマートフォンでスズメのさえずりを録音した。そのさえずりから1フレーズを切り取り、フーリエ変換をするためにテキストファイル化した。この際に奥村晴彦氏のプログラム<sup>[2]</sup>を引用した。

以下に基準とした音声スペクトルと録音した音声スペクトルを示す。X軸が15000[Hz]までとなっているのは以降に特徴が現れなかったためにこの値まで表示している。

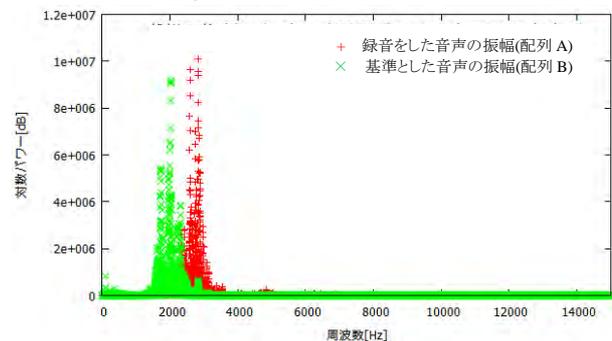


図1 基準と録音した音声データのスペクトル

図1を見ると値のピークが似ていることが分かる。配列 A, B の一致数を数え一定値(今回の値は1050)を超えたため、録音した鳥の種類はスズメとなる。

## 4. 今後の課題

今後の課題は2つある。1つ目はフーリエ変換のアルゴリズムを高速フーリエ変換にすることで判定結果をより高速に求めること。2つ目は本研究で書いたプログラムをAndroidに移植すること。この2つが今後の課題となっている。

## 5. 参考文献

- [1] 荒木有雅弘, “フリーソフトでつくる音声認識システム”2007年  
[2] 奥村晴彦, “WAVE (.WAV) file format”, <http://oku.edu.mie-u.ac.jp/~okumura/wavefmt.html>