

E-077

雑音の平均値を用いた長時間音声からの野鳥の鳴き声の抽出

Detecting bird call from long time sound using noise average

牧野 洋平† 三田 長久† 岩崎 祐介† 高橋 幸司† カムケオ シーパチャン†
 Youhei Makino Nagahisa Mita Yusuke Iwasaki Koji Takahashi Khamkeo Sypachanh

1. はじめに

近年の技術革新に伴う環境破壊に対し科学技術の応用により環境保全に貢献しようという観点から自然環境評価システムの開発を目指す。その一環として環境の変化に敏感な野鳥、特に視認の困難な夜行性鳥類の鳴声による種識別を自動的に行う生息地域等の調査の為のシステムの開発を行っている。

開発段階の現状では予め用意した鳴声の音声データを用い識別試験を行っているが、実用段階では森の中等にレコーダーを設置し、一晩中周りの音を録音した音声を用いることになる。その場合当然録音した物の中に常に鳥の鳴声記録されてはならず、長時間の録音データの中から野鳥の鳴声を抽出する作業も自動化する必要がある。

そこで本研究は長時間音声からの鳴声部分の抽出方法を検討する。具体的には音声波形の分散や自己相関を計算しそこから雑音の平均値を定め、それを音声全体の平均とみなし分散を再計算し平均からのズレが大きい時間帯を鳴声、小さい時間帯を雑音とみなす。加えて雑音を抑制するためのスペクトルサブトラクションを行う。

2. 扱う音声の長さ

実用の際に扱う音声データは一晩中録音を続けたもので長さは数時間に及ぶ。しかしシステムが処理可能な長さの限界は一度全データを読み出す必要があるため通常のCD等の再生よりも短く、使用するマシンの性能にもよるがおよそ10分程度である。さらに入力した音声の長さから処理にかかる時間は正比例せず5、6分を越えた辺りからデータの長さに対する計算時間が大きくなり効率が悪くなるので全体の音声データを5分ずつに分割しシステムに入力する。その5分の中でも収録されている音声には大小の差があるので、入力した5分のデータをシステム内でさらに30秒ずつに分割しその区間の音の大きさに合わせ縮尺を変化させ処理を行う。

また入力する音声データのサンプリング数は44100サンプル/秒だが、鳴声部分の抽出にはそれほど細かい音の性質は必要ないので計算量削減の為に300サンプル/秒まで落とし処理を行い、抽出する時間帯を求めそれに合わせた値を元のサンプリング数のデータから抽出し識別にかける。

3. 雑音の抽出

3.1 雑音の平均値

音声データを聞くだけなら鳴声がかわだてて聞こえる。しかし音声を波形で表すと、雑音の多い音声の場合は鳴声と雑音波形の差が小さく単純に音の大きさだけの抽出

は困難である。加えて音声波形は増減を繰り返す文字通りの波であり大きさの大小を比較すること自体がそのままでは難しい。そのため鳴声の部分を強調する必要がある。

そこで音声 $f(t)$ の各時刻に対して平均、分散、自己相関といった値を求める(式1~3)。

$$\text{平均 } \bar{f}(t) = \frac{1}{30} \sum_{n=-14}^{15} \left| f\left(t + \frac{1}{300}n\right) \right| \quad (1)$$

$$\text{分散 } \sigma(t) = \sqrt{\frac{1}{30} \sum_{n=-14}^{15} \left(\left| f\left(t + \frac{1}{300}n\right) \right| - \bar{f}(t) \right)^2} \quad (2)$$

自己相関

$$C(t) = \frac{1}{30} \sum_{n=-14}^{15} \left| f\left(t + \frac{1}{300}n\right) \right| \left| f\left(t + \frac{1}{300}n + 0.05\right) \right| \quad (3)$$

これらの操作により音声波形から鳴声の部分を浮かび上げらせる(図1)。

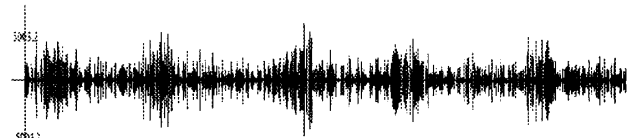


図1.a もとの音声波形



図1.b 平均波形



図1.c 分散波形



図1.d 自己相関波形

平均については雑音との差が小さいため以降は分散と自己相関を用いて考えていく。

こうして鳴声部分を浮上らせた事を利用しそのまま分散、自己相関から大きい値が得られている時間帯を鳴声として抽出することも可能かもしれない。しかし音声データごとに雑音がどの程度の値を表すかの差が大きいため鳴声部分と雑音部分の境界線を引くことが難しい。加えて抽出した部分を識別にかける際に雑音の影響を抑えるため雑音のサンプルが有ったほうが望ましい。そのため分散、自己相関の値から雑音の平均値を求める。

まず、前提条件として処理を行う30秒間には鳴声よりも雑音のほうが長時間録音されていると考える。つまり30秒の中で最もよく出現している分散、自己相関値を主要な雑音の大きさとみなす事とする。そこで分散、自己相関値をいくつかの区間に分割し(図2)、そこから30秒間で

†熊本大学大学院 自然科学研究科

各区間に何回の信号が出現しているかをカウントする。そして最も多く出現した区間を雑音の主要な大きさを表す区間とし、該当区間に出現した信号の和を信号の出現回数で割った雑音の平均値を分散値についての物 \bar{n}_σ と自己相関値についての物 \bar{n}_c それぞれに定める。この時大きい値については鳴声によるものとみなし計算から除外しておき前提とした鳴声よりも雑音のほうが多いという条件から外れている場合でも対応できるようにした。



図2 分散, 自己相関値の区間わけ

3.2 分散値の再計算

ここから分散, 自己相関の分散値を再計算する。ただしこの時分散を計算するための平均値に前項で求めた雑音の平均値を30秒全体での平均値とみなし計算する。分散の大小は平均値からのズレの大小を表すので分散の再計算値が大きい(雑音平均からのズレが大きい)時間帯を鳴声, 小さい(ズレが小さい)時間帯を雑音とみなす。ただし, 分散, 自己相関ともに鳴声のほうが雑音よりも大きな値が得られているので分散の再計算の際には雑音の平均よりも大きい値のみを対象とし, その場合二乗や平方根の計算は必要がないため省略する。このため鳴声探索と雑音探索のための分散値を別々に計算する(式4,5)。

鳴声抽出のための計算

$$\sigma_{2s}(t) = \frac{1}{30} \sum_{n=-14}^{15} \left(\sigma \left(t + \frac{1}{300} n \right) - \bar{n}_\sigma \right) \quad (4)$$

雑音抽出のための計算

$$\sigma_{2n}(t) = \sqrt{\frac{1}{30} \sum_{n=-14}^{15} \left(\left| \sigma \left(t + \frac{1}{300} n \right) - \bar{n}_\sigma \right|^2 \right)} \quad (5)$$

自己相関による計算の場合も同様



図3.a 図1.cの分散再計算



図3.b 図1.dの分散再計算

こうして求めた $\sigma_{2s}(t), \sigma_{2n}(t)$ をもとに抽出を行う。抽出する長さを大概の種に対して一塊の鳴声が収まる範囲として1秒に定め, $\sigma_{2s}(t)$ の t 周辺の1秒分の和

$$\text{sum}_{-\sigma_{2s}}(t) = \sum_{n=0}^{299} \sigma_{2s} \left(t + \frac{1}{300} n \right) \quad (6)$$

が大きい時刻を抽出開始時刻とし, 同様に $\sigma_{2n}(t)$ の和

$$\text{sum}_{-\sigma_{2n}}(t) = \sum_{n=0}^{299} \sigma_{2n} \left(t + \frac{1}{300} n \right) \quad (7)$$

が小さい時刻を雑音の代表の開始時刻とする。

3.3 抽出結果

次に問題になるのは何秒分を抽出するかということである。もちろん音声ごとに抽出すべき鳴声は何秒分含まれているかは異なるので一定の数にすることは出来ない。そのため一番初めに抽出した箇所の分散の再計算値の合計を基準にどこまで抽出を行うかを決定することにし, 分散をもとにして計算した場合は最初に抽出した1秒分の50%まで, 自己相関をもとにした場合は40%までを抽出することとした。ただし, 同じ区間を何度も抽出さぬように一箇所抽出するたびにその区間の値を消去し抽出区間の重複は起こらないようにする。

そのようにして抽出を行った結果が図4のようになる。音声はほとんど雑音の含まれないアオバズクの鳴声にピンクノイズを付加したものをうい横のラインが引かれている部分が抽出区間である(赤:分散, 青:自己相関をもとにした抽出)。

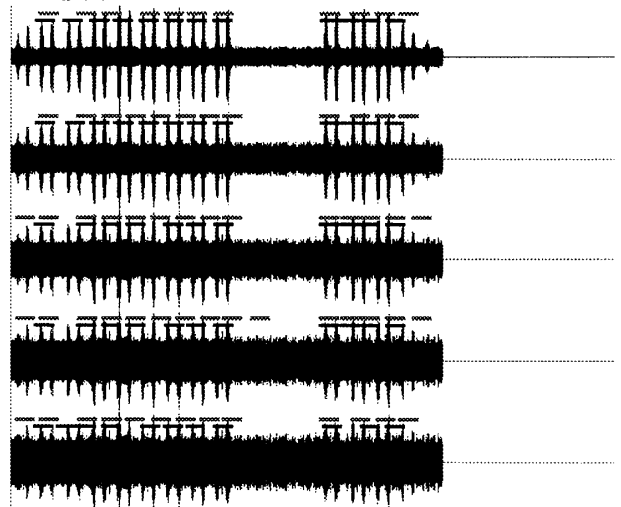


図4 抽出結果

傾向としては分散をもとにして抽出を行った場合は雑音が増えるにつれて雑音に反応して誤った抽出が見られ, 自己相関を元にした場合はその点は安定しているが所々に取りこぼしが見られる, しかしどちらの場合も大まかには鳴声部分を指定できている。

雑音については鳴声の区切りを30秒と短く区切っていることから雑音の変化はそう大きくないものとみなして最初に抽出した一つのみを代表として用い, それをスペクトルサブトラクションに用い雑音抑制を行う。

4. まとめと今後

今回は長時間音声からの鳥の鳴声の自動抽出の方法について検討を行い, 概ね良好な結果が得られた。今後の課題としては作成段階の為に余計な計算が含まれるシステムを簡素化することによりリアルタイム性を高めること, スペクトルサブトラクションの基本手法を用いるに留まっている雑音対策を進めることが挙げられる。

本研究は一部を環境技術開発等推進費の補助を受けて実施した