

ランダムフォレストを用いた シンバルの音の識別における特徴量の重要度可視化

吉川 晃登[†] 竹内 太法^{††} 立蔵 洋介[†]
[†] 静岡大学工学部 ^{††} 静岡大学創造科学技術大学院

1. はじめに

シンバルの音は、叩き方、ドラムスティックの種類、スタンド、部屋など様々なものに影響を受けて変化する。シンバルのような非調波音のスペクトログラムは非常に複雑な構造をしており、シンバルの音の解析は容易なものではない。シンバルの音色に関する研究はあまり行われていない[1]。

本研究では、シンバルの音の解析を簡単化、効率化することを目的とする。機械学習を用いてシンバルの音の違いが存在する特徴を可視化できれば、スペクトログラムを目視で比較する方法より遥かに簡単に解析を行えると考えられる。本研究では、ランダムフォレストを用いたシンバルの音の識別における重要度の可視化を行い、シンバルの音の解析への有用性を検証する。

2. ランダムフォレストによる識別

ランダムフォレストは、複数の決定木を組み合わせたアンサンブル学習アルゴリズムである。決定木は情報利得 $IG(t)$ を最大にするように構築される。 $IG(t)$ は

$$IG(t_p) = I(t_p) - \sum_{j=1}^2 \frac{N_j}{N_p} I(t_j) \quad (1)$$

で表される。ここで、 $I(t_p)$ は親ノード t_p のジニ不純度、 N_p は親ノード t_p におけるサンプル数、 $I(t_j)$ は j 番目の子ノード t_j のジニ不純度、 N_j は j 番目の子ノード t_j におけるサンプル数である。ブートストラップ法と特徴量のランダムサンプリングにより多種多様な決定木を構築し、各決定木の識別結果の多数決をとることで最終的な識別結果を決定する。

ランダムフォレストでの識別では特徴量毎に分類の際の重要度を算出することができる。特徴量 k の重要度を $M(k)$ として、

$$M(k) = \frac{\sum_{i=1}^{n \in F(k)} IG(t_i)}{\sum_{i=1}^{m \in U} IG(t_i)} \quad (2)$$

で定義される。ここで、 $F(k)$ はある特徴量 k が分割対象となるノードの集合、 U はノードの全体集合、 $IG(t_i)$ はあるノード t_i における情報利得である。

3. 重要度のカラーマップ可視化

2 種類のドラムスティックでシンバルのボウを叩いた音をそれぞれ 45 回ずつ計 90 回分収録した自作データセットを用いた。特徴量としてメルスペクトログラムを用いた。サンプルレートを 44100 Hz、メルフィルタバンクのチャンネル数を 128、フレーム長を 46.4 ms、フレームシフトを 46.4 ms、使用したフレーム数を 60 とした。また、結果の信頼性を高め

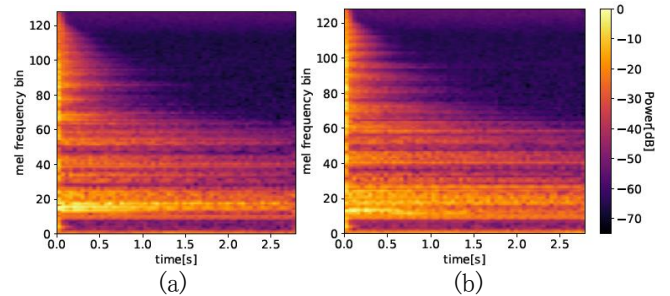


図 1 2 種類のドラムスティックでシンバルのボウを叩いた音のメルスペクトログラム

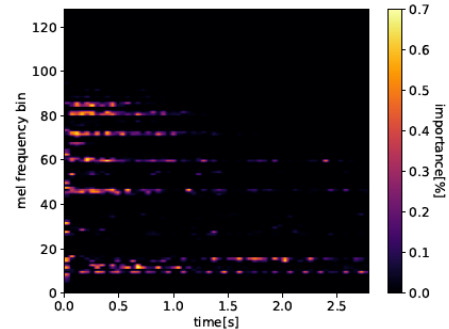


図 2 ランダムフォレストでの識別における特徴量の重要度のカラーマップ可視化

るために 5 分割交差検証を行った。

使用した 2 クラスのシンバルの音のメルスペクトログラムの例をそれぞれ図 1 に示す。ランダムフォレストでの識別における特徴量の重要度を算出し可視化したものを図 2 に示す。重要度が特定の帯域に集中する傾向にあることが確認できる。このことからドラムスティックを変えることによって特定の帯域に違いが生じることが示唆された。さらに、重要度の分布が立ち上がり直後から 0.5 秒までのフレームに偏る傾向にあることも確認できる。立ち上がり直後から 0.5 秒までのフレームに重要度が偏る一つの要因としてシンバルの音の高域成分の減衰が早いことが挙げられる。また、0.5 秒以降徐々に重要度が下がっていく傾向となった。このことからシンバルの減衰の仕方にばらつきがあったことが示唆された。

今回用いた 2 クラスのシンバルの音について比較解析を行う際には、主に重要度の偏っている 0.5 秒以前のフレームの重要度の高い帯域に着目すべきであることが示唆された。今後は重要度可視化結果の定量的な評価および加工したデータを用いた主観評価実験を行いたい。

参考文献

[1] K. Werner, *et al.*, Proc. ASA, vol. 25, no. 1, p. 035004, 2015.