

環境音とその声真似の分析

D-14

Analysis of environmental sound and voice imitation

渡邊 貴波[†] 小坂 直敏[†]Tattowa WATANABE[†] Naotoshi OSAKA[†]

† 東京電機大学大学院未来科学研究科情報メディア学専攻

† Tokyo Denki University

1 はじめに

効果音は、映画やアニメーションなどのコンテンツにおいて重要な役割を果たしている。制作者は、場面毎に適した環境音を選択するために、環境音データベースによる検索を用いる。環境音検索はテキストによる検索が主であるが、テキストだけでは、多面的な音のニュアンスを表現することは難しい。そこで、われわれは声真似入力により環境音を検索する方法を進めている。先行研究で、Blancasが声真似入力により、音データベースの中から環境音を検索するシステムを提案している。このシステムは声真似の識別辞書を作成して、入力の声真似の種類を分類して検索に用いている[1]。

類似した技術である音声認識では、話者が異なると、話者適応技術が必要である。一方、声真似による環境音検索は、異なる話者間の音声よりも、さらにスペクトル特徴量の距離があると考えられる。

本稿では検索の前段階として、どのような種類の環境音と声真似が類似しているか調べる。そのために、環境音とその声真似に対し、時間の非線形伸縮をするDPマッチング[2]を行い、スペクトル情報を表すMFCCを用いて2音の距離を求めて、分析した結果を述べる。

2 分析に用いた環境音と声真似

分析に用いたデータは環境音100個と声真似100個である。それらの音を以下のように分類した。

- ①「単体音, 繰り返しなし, 減衰系」は、図1の①のように音量が立ち上がって、すぐに減衰する音である。
- ②「単体音, 繰り返しなし, 定常系」は、図1の②のように音量が一定区間定常な音である。
- ③「単体音, 繰り返しあり」は、図1の③のように同じ単体の音が複数ある音である。
- ④「連続音」は、図1の④のように複雑な山の音である。

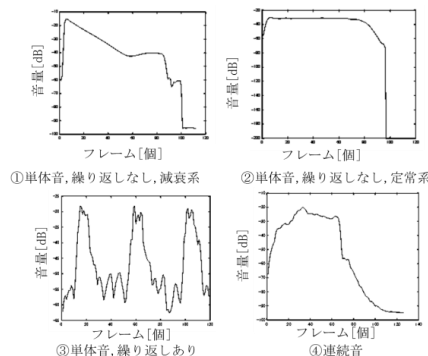


図1 分類した4種類の環境音のパワーのグラフ

3 分析結果と考察

環境音とその声真似のそれぞれに対しMFCCの距離を求め、DPマッチングにより、距離が近い順にランク付けし、上位の音と下位の音のソナグラムを分析した。図2上側は、プロペラ機の飛ぶ音とその声真似のソナグラムで、ランク上位に属する。下側は、猫の鳴き声とその声真似のソナグラムで下位に属する。プロペラ機の飛ぶ音と声真似は、ソナグラムの模様が類似している。これは、元の音は調波構造を持たない比較的定常な雑音であり、声真似が容易であったためと考えられる。一方、猫の鳴き声と声真似は、ソナグラムの模様が異なる。すなわち、猫の声帯振動形状を声真似では再現できないためと考えられる。

上位の音と下位の音のソナグラムを分析した結果、スペクトルの変化が単純な音は、声真似との距離が近いことが分かった。また、調波構造が全く異なる動物の声は声真似音声と距離があることが分かった。

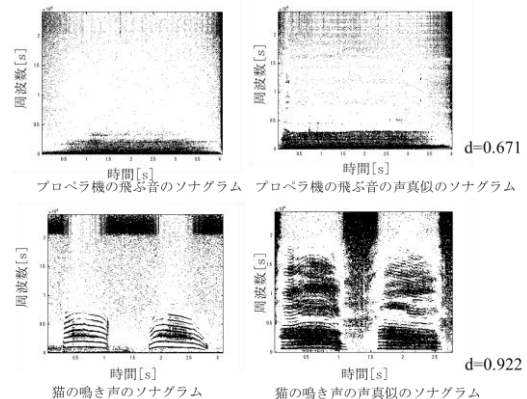


図2 上位の音と下位の音のソナグラム

4 まとめと今後の展望

環境音をその声真似から検索する問題で、2音のMFCCを用いた距離の大小を調べ、その原因を考察した。その結果、MFCCによる距離だけでは2音の距離を小さくするのは困難であると分かった。今後は他の特徴量を用いることを検討する。

参考文献

- [1]Blancas, et al, "Sound retrieval from voice imitation queries in collaborative databases." *AESC: 53rd International Conference: Semantic Audio*, 2014.
- [2]古井貞熙, "新音響・音声学", pp. 184-186" 東海大学出版会, ISBN4-486-00896-0, 1985年 第1版