

# 学内で観測される環境音の分析とその識別

愛知工科大学 工学部 情報メディア学科 中西遥輝

## あらまし

その場の状況理解のために、学内の様々な場所での環境音を観測し、識別精度を検討する。愛知工科大学内の食堂やバス停前などの4カ所において観測される環境音を対象とする。評価として、場所ごとに観測された環境音に対する識別と、観測された全ての環境音に対する識別を行う。実際の環境で環境音を収録し、Deep Neural Network (DNN)で学習、識別精度を評価する。また、観測された環境音を使用し、場所自体の推定を行う。実験により、場所別での環境音の識別では、正解率がそれぞれ、食堂93.93%、バス停前93.54%、短大生ホール93.91%、グラウンド93.40%を得た。全ての環境音では92.87%を得た。場所推定では93.55%と比較的高い正解率を得ることが出来た。

### 1. はじめに

通信端末やロボットに関する技術が成長を遂げてきた現代、多くの場面で音声対話技術が活用されている。しかし、このような技術も周囲の環境によって認識の精度が落ちてしまうことがある。人が混みやすい場所では、正確な認識を行うことが出来ず間違った返答をしてしまう場合がある。環境音に関する研究は、文献[1][2]などが挙げられるが、複数カ所の周囲の状況を把握するための検討はほとんど行われていない。

そこで、愛知工科大学内の様々な場所で観測される環境音を識別し、その場所の状況を把握すると共に観測された環境音データを用いて場所の推定精度を検討する。

### 2. 学内施設での環境音の収録

食堂やバス停など学内施設4カ所で観測される環境音を収録した。平日授業時間の11時00分～18時00分の時間帯で収録を行った。観測された環境音からは、それぞれ、食堂15種類、バス停前18種類、短大生ホール13種類、グラウンド14種類の環境音を得られた。また、その場所でのみ観測される環境音はすべて合わせて24種類であった。

### 3. 識別手法

収録された環境音を人手によってラベリングを行い、DNN(Deep Neural Network)により識別実験と場所の推定を行う。流れとしては、以下の通りである。

- ① 録音データを分析し、特徴量として1フレーム当たり40次元メル周波数スペクトルを抽出。
- ② 前後の特徴量を考慮するため、各フレームの前後5フレームを1つの特徴量ベクトルとして再構築。
- ③ DNNを用いて、ニューロン数100、200、300で環境音の識別実験と場所を推定。

### 4. 実験

場所ごとに観測された環境音と観測された全ての環境音の2パターンで識別実験を行った。図1に示すように、場所ごとに観測された環境音の場合、どの条件でも全体的に93%台の正解率であった。観測された全ての環境音の場合、どの条件でも92.8%台であった。2パターンとも大きな変化は見られず、全体的に高い正解率となった。

場所の推定は、その場所のみで観測される環境音のみが検出された状態で実験を行った。図2に示すように、最も正解率が高かったのは、隠れ層4層、ニューロン数

100で93.55%であった。実験結果からその場所のみで観測される環境音を頼りに場所の推定を行うことが現状可能と考えられる。

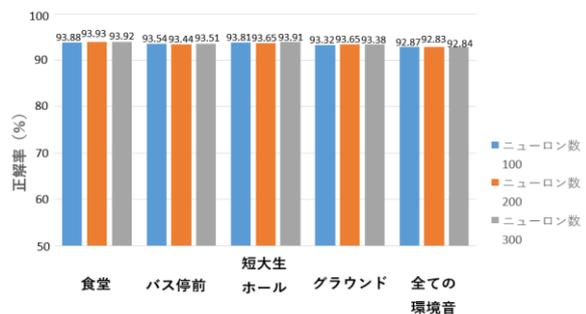


図1. 環境音の識別実験の実験結果

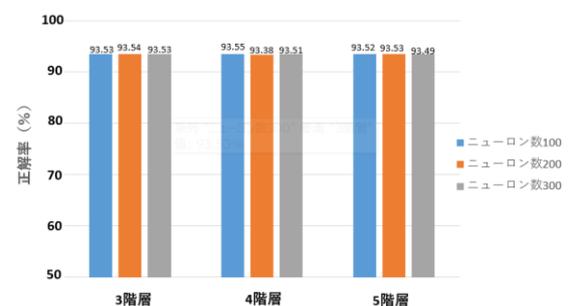


図2. 場所の推定の正解率

### 5. まとめ

本研究では、学内施設4カ所で観測される環境音の識別と観測された環境音を使用した場所の推定の検討を行った。場所ごとに観測された環境音の場合、93%台の正解率を得た。観測された全ての環境音の場合92.87%であった。場所の推定の実験を行った結果、93.55%と比較的高い数値であった。今後の課題は、データ量を増やしてより精密な正解率を出すこと、場所の推定に関しては更なる検討が必要である。

### 参考文献

- [1] 三木ら, 電気電子情報通信学会技術者研究報告, NLC99-38, SP9-106, pp79-84, 1999年
- [2] 都築亮登, 2020年