

# ニューラルネットワークを用いた室内アラーム音の識別の検討

門倉 丈<sup>†</sup> 渡辺 滉平<sup>†</sup> 田中 博<sup>†</sup>

神奈川工科大学 大学院 情報工学専攻<sup>†</sup> 神奈川工科大学 情報学部 情報工学科<sup>†</sup>

Elisa sihombing<sup>‡</sup> Syauqan Wafiqi<sup>‡</sup>

Electronic Engineering Polytechnic Institute of Surabaya<sup>‡</sup>

## 1. まえがき

室内で発生する様々なアラーム音を、聴覚障害を持つ人や、加齢により耳が遠くなった人が聞き分け、識別することは困難である。そのため本稿ではニューラルネットワークを用い、室内で発生するアラーム音からその音源を識別することを目的とする。今回は8種類の異なる音源を用意し、それぞれから発生する音を録音し検討に利用する。学習に使用する特徴量は、周波数ごとの強度を示すパワースペクトルとメル周波数ケプストラム係数(MFCC)である。2種類の特徴量を用いた実験結果を比較し、識別精度の評価を行う。

## 2. 特徴量の抽出

本検討では、一般家庭に配備されていると思われる8種の警報器を用いて実験を行った。スペクトログラムの一例(玄関ベル、ガス警報器)を図1に示す。周波数特性に違いがあることが確認でき、音源の識別が可能であると考えられる。今回の録音条件は、サンプリング周波数16,000Hz、時間は60秒、モノラル録音である。STFTを用いフレーム長32ms、フレームのシフト幅を10msとし8,000Hzまでのパワースペクトルとして特徴量を抽出した<sup>[1]</sup>。音の大小による影響は、得られたスペクトルの最大値で正規化を行うことで排除した。

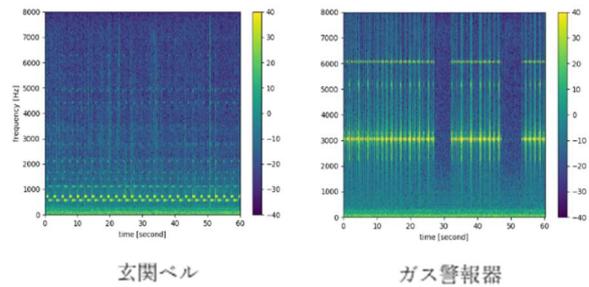


図1 アラーム音のスペクトログラム

MFCCの抽出はフレーム長、フレームのシフト幅はパワースペクトルと同一とし、メルフィルタバンクのチャンネル数20、次元数は1フレームにつき13とした。3フレームで1つの特徴量とし39次元の特徴量を作成した<sup>[2]</sup>。

## 3. ネットワークの構造

本実験で使用したニューラルネットワーク<sup>[2]</sup>を図2に示す。中間層を2層持つ単純なネットワークであり、過学習を防止するために中間層のユニットを50パーセントの確率で消失させるドロップアウト法を用いている。パワースペクトルを用いる場合と、MFCCを用いる場合の実験において両者の違いは入力層のユニット数だけである。

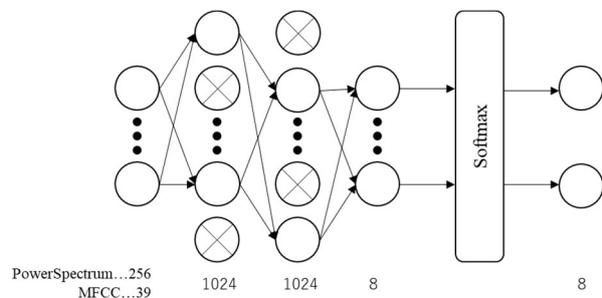


図2 ネットワークの構造

“Classification of indoor alarm sound using neural network”

Takeru Kadokura<sup>†</sup>, Kohei Watanabe<sup>†</sup>, Hiroshi Tanaka<sup>†</sup>

Kanagawa Institute of Technology<sup>†</sup>

Elisa sihombing<sup>‡</sup>, Syauqan Wafiqi<sup>‡</sup>

Electronic Engineering Polytechnic

Institute of Surabaya<sup>‡</sup>

4. 学習モデルの作成

本実験では学習に 60 秒×2 個のデータを用いた。以下にパワースペクトル, MFCC を用いた学習曲線をそれぞれ示す。損失関数には交差エントロピー誤差を用い, ミニバッチのサイズは 20 とした。ミニバッチに対する精度 (Accuracy) と誤差関数から求められる誤差 (Loss) の収束を確認したため今回の Epoch は 1000 回とした。

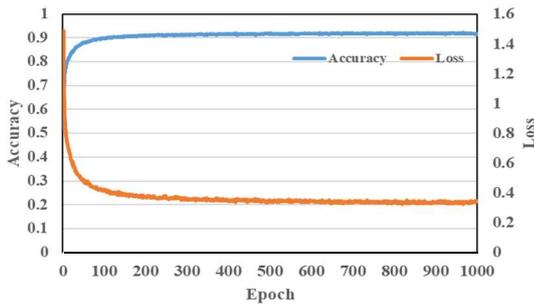


図3 パワースペクトルを用いた学習

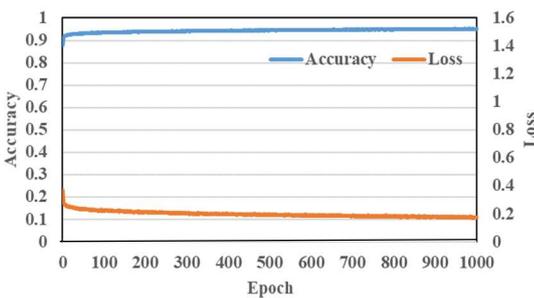


図4 MFCCを用いた学習

5. 識別精度の評価

実際にシステムに本手法を利用する場合, 短時間での識別が必要となるため, 識別精度の評価実験は各警報器に対して 5 秒×10 個のデータを用いて行った。識別に用いた警報器を表 1 に示す。録音条件は学習用データを作成した時と同一である。識別精度を表 1, 表 2 に示す。パワースペクトルは高い精度での識別ができているが, MFCC では目覚まし時計 2 の識別ができず十分な精度とはいえない。

表1 使用した警報器

A:ドア開放警報器	B:火災報知器 1	C:火災報知器 2
D:ガス警報器	E:玄関ベル	F:ヤカン
G:目覚まし時計 1	H:目覚まし時計 2	

表2 パワースペクトルを用いた識別結果

	A	B	C	D	E	F	G	H
A	9	0	0	0	1	0	0	0
B	0	10	0	0	0	0	0	0
C	0	0	10	0	0	0	0	0
D	0	0	0	10	0	0	0	0
E	0	0	0	0	10	0	0	0
F	0	0	0	0	0	10	0	0
G	0	0	0	0	0	0	10	0
H	0	0	0	0	0	0	0	10

表3 MFCCを用いた識別結果

	A	B	C	D	E	F	G	H
A	9	0	0	0	1	0	0	0
B	0	10	0	0	0	0	0	0
C	0	0	10	0	0	0	0	0
D	0	0	0	10	0	0	0	0
E	0	0	0	0	10	0	0	0
F	0	0	0	0	0	10	0	0
G	0	0	0	0	0	0	10	0
H	0	0	0	0	0	0	10	0

6. まとめ

室内で発生するアラーム音をニューラルネットワークを用いて識別した。その際の特徴量にパワースペクトルとメル周波数ケプストラム係数 (MFCC) の 2 種類を用い, 両者の精度を比較した。結果からパワースペクトルは短時間のデータでの識別が可能であったが, MFCC では識別に誤りが生じた。実用的なシステムとして運用する場合, 特徴量にはパワースペクトルを用いた学習モデルが有効である考えられる。また, 識別対象に室内で発生するアラーム音を想定していることから, 様々な環境音の影響が考えられる。それらの音がアラーム音に重畳した場合の評価が必要である。

参考文献

[1]牧野浩二, 西崎博光, インタフェース, Vol.8, pp.36-40, 2017  
 [2]牧野浩二, 西崎博光, 算数&ラズパイから始めるディープラーニング, pp.78-81, 2018