ISS-P-020

スペクトログラム画像と GoogLeNet を用いたアラーム音の識別実験

藍原　直大†　　佐野　将太††　　田中　博†
† 神奈川工科大学　情報学部　情報工学科
†† 神奈川工科大学大学院　情報工学専攻

1. はじめに

音を画像に変換して，画像処理技術を適用することで雑音除去や音声識別を行う研究が行われている．[1] 本稿では，聴覚障がい者支援等を目的としたアラーム音識別において，従来の手法[2]より良好な結果が得られたので報告する．

2. 入力データの作成

本実験では図 1 のように，音データを STFT（短時間フーリエ変換）によってスペクトログラム画像に変換する．後述の学習モデルの入力データとするために，各切り出し秒数で区切り，画像を正方形化した．学習する 7 種のアラーム機器を表 1 に示す．アラーム音が出力されない場合も識別するために，日常空間を想定して，エアコンと扇風機が動作している部屋の環境音，更に人の話し声が入った会話音の計 2 クラスを追加した．

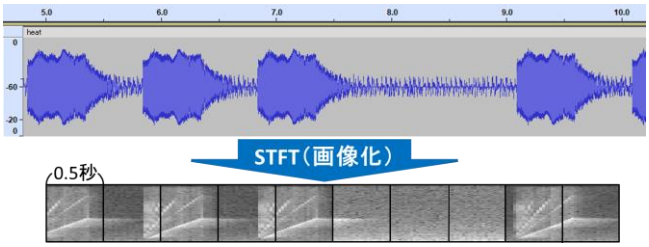

図 1　波形の画像化

表 1　識別対象の音源

窓アラーム	煙感知器	熱感知器
ガス漏れ報知器	インターホン	キッチンタイマー
目覚まし時計	環境音	会話音

3. 識別モデルの作成

識別のための学習モデル作成には，必要となるデータ数や学習時間から，事前学習モデルの GoogLeNet を適用した．各学習パラメータを表 2 に示す．実際のアラーム通知を想定し，識別にかかる時間を変化させ精度と比較する．図 2 に示した画像のように，切り出し秒数を変えてそれぞれ学習し，4 つの識別モデルを作成した．学習用データは1つのモデル当たり 270 枚で行った．

表 2　学習パラメータ

切り出し秒数	バッチサイズ	エポック	学習率
0.5	4	400	1e-5
1，2，3	4	100	1e-5

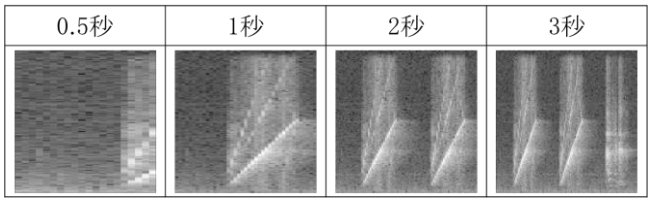

図 2　切り出し秒数ごとのスペクトログラム画像

4. 識別結果

作成した各識別モデルを，学習に用いていないテストデータ 180 枚で評価した．切り出し1秒のモデルは精度 98.89%（180中 2 誤判定），2秒は 98.33%（180中 3 誤判定），3 秒は 100.00%と，極めて高精度な結果が得られた．切り出し 0.5 秒のモデルは 95.56%（180中 8 誤判定）で，各クラスの混合行列を図 3 に示す．主に熱感知器と煙感知器を会話音と誤判定していた．この 2 種の音源はアラーム音の間隔が大きく，最大で 1.6 秒ほど空いている．図 1 に示すように，短い切り出しでは識別に十分な特徴が出ていない画像が生成されていた．一方，音が出力されている時間の画像は正しく判定していたことから，短い識別時間でも高精度な識別は可能であると考える．

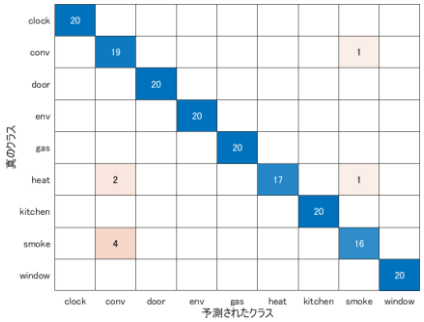

図 3　切り出し 0.5 秒の混合行列

5. まとめと今後の課題

GoogLeNet を用いてアラーム音を短時間に高い精度で，識別することができた．今後はリアルタイムでの識別実験や，ワンボードコンピュータへの組み込みとその応用を考える．

参考文献

[1] 佐野将太他，“スペクトログラム画像を用いた室内音環境の識別手法とその実環境評価，” HCG シンポジウム，HCG2022-B-1-1，6pages，2022.

[2] 門倉丈他，“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とその報知システムの基本検討，” 情報システム研究会，IS-19-018，pp. 91-96，2019.