

集音器を用いた音の画像認識

Sound image recognition by Edge AI using a sound collector

梶本 阿斗

霜山 竜一

Ato Kajimoto

Ryuichi Shimoyama

日本大学大学院生産工学研究科

Graduate School of Industrial Technology, Nihon University

1. まえがき

聴覚障害者は外出時に側方や後方から接近する自動車やバイクに気づかず危険にさらされる場合がある。著者らは音源の位置を推定し障害者に呈示するウェアラブル聴覚支援システムに、風などの環境音による影響を抑える目的で Edge AI を導入した¹⁾。

本報告では、同システムに新たに集音器を導入することで、雑音を抑制して遠方の音が検出できるかどうかについて検討した。

2. GCP と USB Accelerator について

Google Cloud Platform(GCP)の AutoML Vision API を使用して環境音を画像として認識させた。AutoML Vision API はプログラムの起動に時間を要するうえ、常時接続すると利用料金が高額になるためリアルタイム処理には適さない。GCP で開発した API を USB Accelerator(Coral) 付きの PC で動作させ、ローカル環境で音を認識させた。

3. イヤマイクと首掛け集音器

音の検出センサとしてイヤマイク(MDR-EX31BN, SONY)と首掛け集音器(SMR-10, SONY)の2種類を採用して実験結果を比較した。ノイズキャンセリングインナーイヤホンの雑音除去用の外部マイクをイヤマイクとして使用した。首掛け集音器はテレビのスピーカや補聴器として使用可能で周囲の雑音をキャンセルする機能があり、個々の使用者の聴覚特性に合わせて聴こえ方が調節できる。集音器に付属されたインナーイヤホンの信号出力を利用した。この集音器を用いれば雑音を抑えて遠方の音が検出可能になる。

4. 処理の流れ

AutoML Vision API を用いて環境音を認識するための処理の流れを図1に示す。インターネットや実際に録音して得た音響波形の wave ファイルを周波数スペクトログラムに変換した(図2)。測定時間を1[s]、測定周波数帯域を500[Hz]~15[kHz]に設定した。音のレベル差が画像に及ぼす影響を抑制するため音のレベル

を規格化して配色した。また、遠くまで伝搬しやすい、低い周波数の音の成分を強調するために周波数軸を対数表示した。この周波数スペクトログラムを画像とみなして9ラベルに分類した画像を AutoML Vision に学習させた。マイクロホンで検出した音を wav ファイル化した後に画像化して、学習済み AI に認識させた。各ラベルの認識率を求め最も認識率の高いラベル名を出力した。図3に聴覚支援システムと Edge AI の構成図を示す。イヤマイクと首掛け集音器のいずれかを用いて音の認識率を比較した。Edge AI の出力信号は聴覚支援システムを動作させるトリガとして使用する。

学習用画像はインターネット等から取得した wave ファイルから作成した。学習用の画像の例を図4に示す。

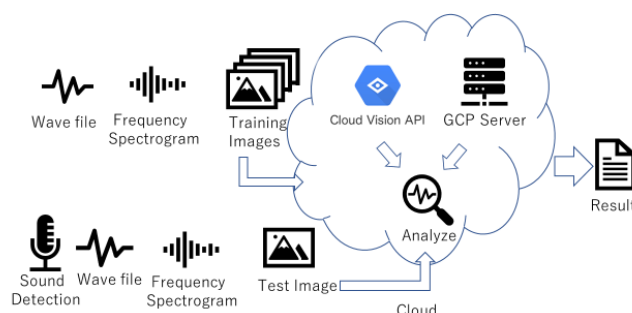


図1 処理の流れ
(AutoML Vision API)

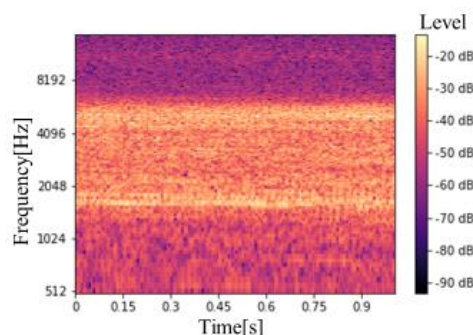


図2 周波数スペクトログラム(自動二輪車)

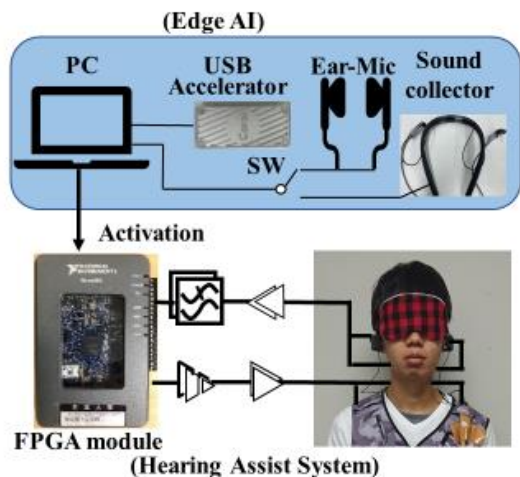


図3 聴覚支援システムと Edge AI の構成図

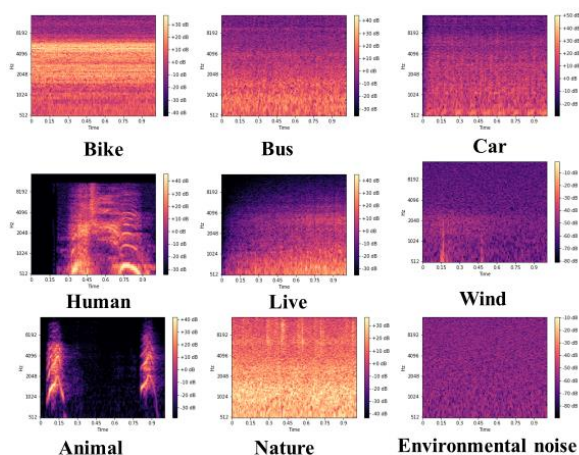


図4 学習用画像(9 ラベル)

学習させた音は自動二輪車の走行音、バスの走行音、車のエンジン音、人間の笑い声、生活歩行音、動物の鳴き声、雨の音、環境雑音、風の音である。ラベル名は自動二輪車、バス、普通乗用車、自然音、生活音、動物音、人の声、風の音、環境雑音の9種類である。学習させた画像数は自動二輪車が200枚、バス200枚、普通乗用車200枚、自然音100枚、生活音100枚、動物の音100枚、人の声100枚、風の音100枚、環境雑音100枚の計1200枚である。

4.実験結果

屋外の一般道路脇の歩道で、自動二輪車の走行音を Edge AI で認識させた。イヤマイクと首掛け集音器のいずれかを用いて Edge AI による音の認識率を比較した。計測時の自動

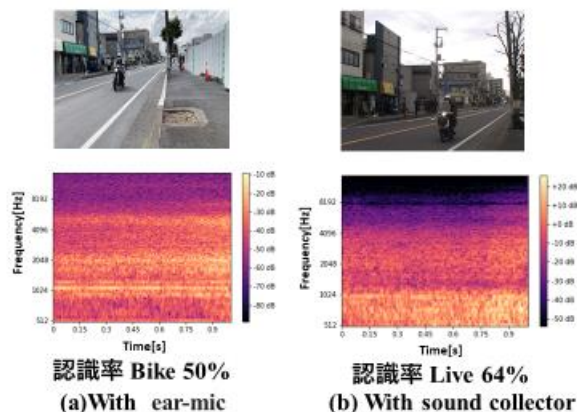


図5 イヤマイクと集音器の認識結果の違い

二輪車の写真と音の画像を図5に示す。同図(a)はイヤマイクを用いた場合、同図(b)は集音器を用いた場合の結果である。図4に示した自動二輪車の学習用画像と比較すると、イヤマイクによる画像(図5(a))はほぼ一致するが、集音器を用いた画像(図5(b))の特徴に著しい違いがみられた。イヤマイクを用いた場合は認識率 50[%]で自動二輪車と認識されたが、集音器を用いた場合に生活音として誤って認識された。補聴器としても使用できる集音器は、人の聴覚特性に合わせて出力信号の周波数特性が調整されていると考えられる。一方、イヤマイクを用いて検出した音や学習用に使用した音の波形はそのような処理がなされていない。これが画像の特徴に違いが生じた理由である。

5.おわりに

イヤマイクを用いた場合に Edge AI は認識率 50[%]で自動二輪車と認識した。集音器を用いた場合には誤って認識された。集音器では人の聴覚特性に合わせて信号出力の周波数特性が調整されている。一方、イヤマイクを用いて検出された音や学習用に使用した音の波形はそのような処理がなされていない。学習用画像とのミスマッチが原因と考えられる。

今後は、Edge AI と聴覚支援システムを連携動作させる予定である。

参考文献

- 1) 梶本阿斗ら：知的聴覚支援システムに関する研究, 電子情報通信学会 2021 年総合大会,2021-8