

## Cloud-base AI による走行するバイク音の認識

日大生産工(院) ○梶本 阿斗 日大生産工 霜山竜一

### 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。著者らが開発した、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<sup>1)</sup>は、屋外では風などの他の環境音に反応するため、動作が不安定になる。音を認識させ特定の音源のみに反応して聴覚支援を行わせる必要がある。環境音をAIで認識させる試みの多くは室内でなされており<sup>2,3)</sup>、屋外で遠方から走行する自動二輪車の音をAIで認識させた例はない。

本報告では、Cloud-base AIを用いて30[m]遠方から走行させた自動二輪車の音を認識/識別可能なことを示す。

### 2. Cloud-base AI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。Google Cloud Platformは、GoogleがCloud上で提供するサービス群の総称である。Cloud上で好きなハードウェアとソフトウェアを選択し、AIのプログラム開発が行える。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。

### 3. 処理の流れ

Cloud-base AIを用いて環境音の認識を行う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。この周波数スペクトログラムを画像とみなして7ラベルに分類したものをCloud Vision APIに学習させた。入力したテスト画像を、Cloud上の学習済みAIに認識させ、各ラベルの認識率を求め最も認識率の高いラベルを選ばせた。屋外で実際に録音した、遠方から走行する自動二輪車の音をAIに認識/識別させた。

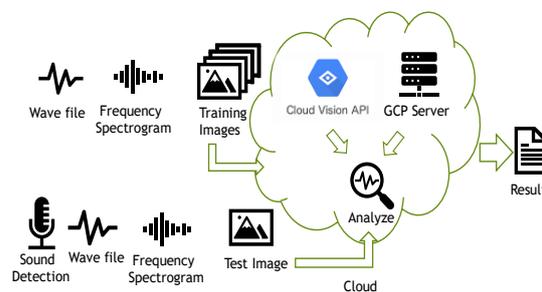


図1 処理の流れ  
(Google Cloud Platform)

### 4. 実験結果

音の信号波形を周波数分析して周波数スペクトログラムを求めた。測定時間を1[s]、測定周波数を20[Hz]~15[kHz]とした。音のレベルによる画像への影響を除く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。周波数スペクトログラムの画像の例を図2に示す。

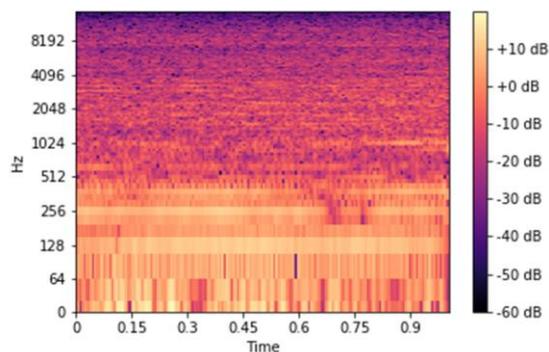


図2 周波数スペクトログラム

学習用画像を図3に示す。学習させた音のラベルは自動二輪車、バス、普通乗用車、自然音、生活音、動物音、人間の声の7種類である。学習に用いた画像数は自動二輪車が300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、人間の声100枚の計1100枚である。

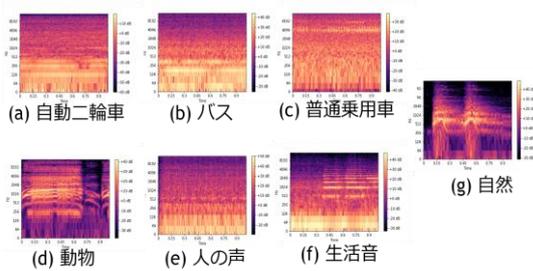


図3 学習させた画像(7ラベル)

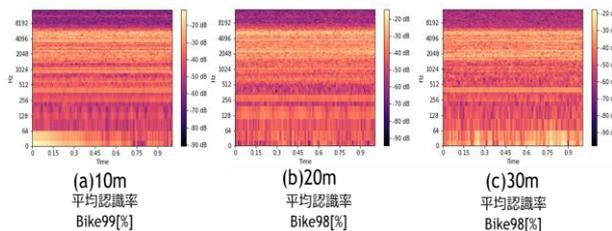


図4 遠方に停止した自動二輪車の音の平均認識率

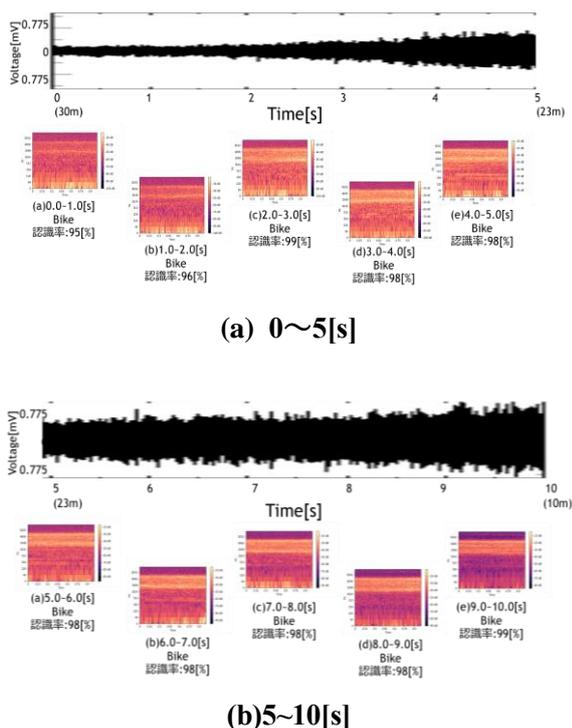


図5 走行する自動二輪車の音の認識結果。

屋外でマイクロホンからそれぞれ 10[m]、20[m]、30[m]離れた位置に自動二輪車を停めてエンジンをふかした音を 10[s]間録音した。

得られた周波数スペクトログラムと AI による平均認識率を図4に示す。録音された波形を 1[s]毎に 10 回分認識させてそれらの平均を平均認識率とした。停止した自動二輪車のエンジン音は、その位置に拘わらず高い認識率 98[%]を示し、7つのラベル中で自動二輪車が最も認識率が高かった。マイクロホンから 30[m]離れた位置から 10[m]手前まで自動二輪車を走行させた場合について検討した。走行時の自動二輪車のエンジン音の時間波形と 1[s]毎の認識率を図5に示す。同図(a)は前半の 5[s]、(b)は後半の 5[s]を示す。音の波形は自動二輪車が接近するにつれて、振幅が大きくなるのがわかる。周波数スペクトログラムは自動二輪車が接近してもパターンや色調があまり変化しないことがわかる。最も遠方 30[m]にある場合でも自動二輪車の認識率は 95[%]で他のラベルより高くなった。AI によって自動二輪車の音を識別できることがわかる。

## 5. まとめ

Cloud-base AIによる環境音の認識について検討した。音の信号波形を周波数分析して周波数スペクトログラムを求めた。周波数スペクトログラムを画像とみなし画像認識用 AIを使用して音の認識を行った。音のレベルを規格化し、周波数軸を対数表示することで、30[m]遠方から10[m]まで接近するバイクの走行音を認識率98[%]以上で認識できた。今後は、Acceleratorを導入してローカルな環境でAIによる音のリアルタイム認識を行う予定である。

## 参考文献

- 1) 関淳 “平成 31 年度修士論文 音源の動きを呈示する聴覚支援システムに関する研究”,序章,研究と背景,pp2
- 2) 水野智喜、田胡和哉著 “音声認識を利用した介護予防・日常生活支援総合事業改善の試み “
- 3) 石原一志、坪田康、奥野博著 “日本語の擬音語表現に着目した環境音の自動認識”