M-020

スペクトログラムを用いた低 SN 環境下における EV 車両音の識別に関する基本検討

Fundamental Study on Classification of EV Sounds in Low SN Environment Using Spectrogram

秋草 直世† 松尾 空† 田中 博† 宮崎 剛†
Naose Akikusa Sora Matsuo Hiroshi Tanaka Tsuyoshi Miyazaki

1. はじめに

電気自動車（EV）やハイブリッド車（HV）の走行音は小さいため、歩行者が気づかず事故のリスクが高まるという課題が指摘されている。この対策として、AVAS（Acoustic Vehicle Alerting System）と呼ばれる警告音装置の搭載が義務化されている。AVAS 音には、500Hz～5kHz の可聴帯域にわたる音を中心とした特徴的な周波数成分を有している[1]。しかし、環境雑音が大きい場面では、AVAS 音が周囲の音に埋もれ、歩行者が認知しにくくなる可能性がある。

筆者らはこれまで、複数の異なる音量の音が混在する環境下での音源識別手法に取り組み、その有効性を明らかにしてきた[2]。本研究では、この識別手法を適用し、環境雑音内の AVAS 音の有無を判定する方法について検討した結果を述べる。

2. 音の収録

2.1 AVAS 音の取得

AVAS 音の取得には日産サクラ（以下、サクラ）および日産ノート e-power（以下、ノート）を使用した。いずれの車種も低速走行時またはシフトレバーを R レンジに入れた際に AVAS が作動するよう設計されている。

本検討では、モーターの回転音やタイヤと道路による摩擦音の影響を排除するため、車両を停止した状態でシフトレバーを R に入力し、その際に発生する AVAS 音を収録した。収録の際は、車両前方 1 m の位置に設置したスマートフォン（iPhone14）の内蔵マイク、収録ソフトは PCM 録音を使用した（図 1）。音声データは、サンプリング周波数 44.1kHz、量子化ビット数 16bit の PCM 形式、モノラルで収録した。

収録した AVAS 音を 16kHz にリサンプリングし、表 1 に示すパラメータでスペクトログラムに変換した画像を図 2 に示す。図より、AVAS 音の仕様に準じて 500～5kHz の範囲に明瞭な周波数成分があることを確認した。

図 1　AVAS 音取得時の状況

表 1　スペクトログラムの変換パラメータ

パラメータ項目	値
窓幅	256
窓の移動幅	128
窓関数の長さ	256
窓関数	Hann
画像 1 枚当たりの時間長（秒）	1.00
画像 1 枚当たりの大きさ（幅×高さ）	124×129

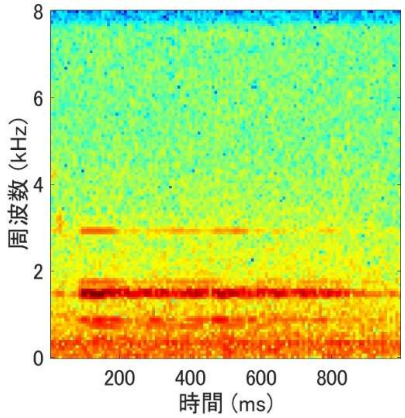

図 2　AVAS 音のスペクトログラム

2.2 環境音

環境音には ATR 環境音データベースを用いた。音源として「バスターミナル」（以下、雑音 A）、および「駅前待ち合わせ広場」（以下、雑音 B）を選定した。音源の選定にあたっては、識別手法の使用が想定される環境で実際の雑踏に近い音響特性を有していること、および両音源に含まれる周波数帯域におけるパワー分布が異なっていることを考慮した。

雑音 A および雑音 B の音声ファイルは、いずれもサンプリング周波数 48kHz、量子化ビット数 16bit、ステレオで収録されている。2 つの環境音にモノラル化処理を行ったうえで 16kHz にリサンプリングし、表 1 のパラメータに基づいてスペクトログラムに変換した画像を図 3 に示す。図より、雑音 A は主に 0～1kHz の範囲に強い周波数成分を持つ一方、雑音 B は 0～5kHz にわたって広く分布する強い周波数成分を有していることが分かった。

3. 識別モデルの学習と結果

本検討では、あらかじめ学習した識別モデルを活用し、スペクトログラムを識別対象とした転移学習を実施する。本章では、転移学習に用いるスペクトログラムデータセッ

†神奈川工科大学情報学部情報工学科　Dept. of Information and Computer Sciences, Kanagawa Institute of Technology

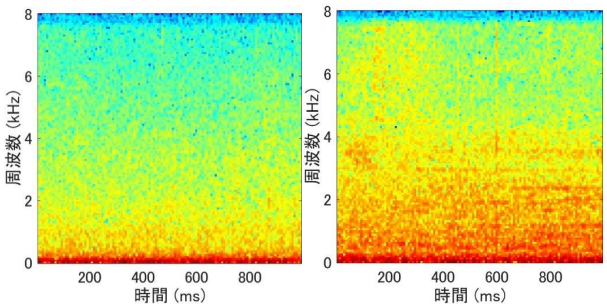

図 3　環境音のスペクトログラム

トを作成するための具体的な手順および識別結果について述べる。

3.1 雑音混入音の作成

サクラおよびノートの AVAS 音を所定の SN 比（0dB，-6dB，-12dB，-18dB，-24dB）の 5 条件となるように音圧を調整し、雑音 A と重畳し、複合音 A を作成した。これにより、元の雑音単体と 5 種の SN 比条件で重畳された音の計 6 クラスの音声データとした。同様の処理を雑音 B に対しても実施し、複合音 B と雑音 B の 6 クラス分の音声データとした。設定した SN 比はデシベル変換式より、隣接する SN 比間でパワー比が 1/4 となる関係にある。

重畳したデータの一例として、表 1 のパラメータを用いて求めた複合音 B のスペクトログラムを図 4 に示す。SN 比の低下に伴って AVAS 音の周波数成分が雑音に埋もれる様子が確認できる。

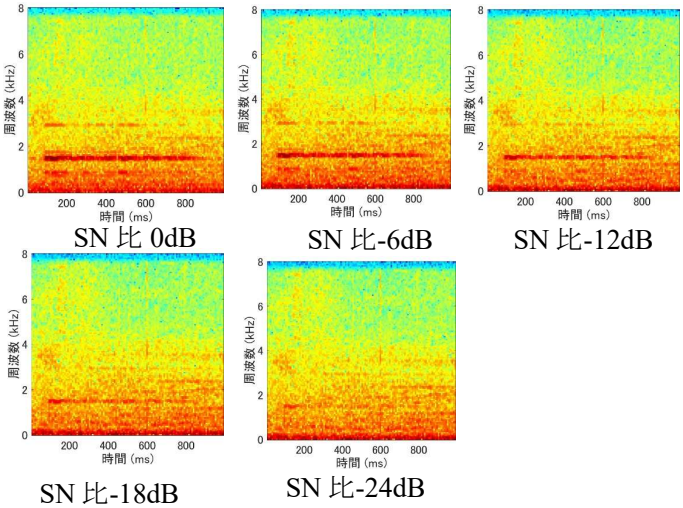

図 4　SN 比ごとのスペクトログラム

3.2 転移学習

複合音 A のうち、サクラとノートのそれぞれ 220 秒分のデータを表 1 のパラメータを用いて、1 クラスあたり 440 枚の画像データスペクトログラムへと変換した。各クラスのラベルは 0dB，-6dB，-12dB，-18dB，-24dB と設定した。

次に、雑音 A の中から 440 秒分を切り出し、同様に 440 枚のスペクトログラムへ変換した。このデータには noise というラベルを付与した。これらのデータをデータセット A とする。

作成した画像データは、クラスごとに学習用、検証用、評価用として、それぞれ 8：1：1 の比率でランダムに分割した（表 2）。これと同様の処理を複合音 B および雑音 B に対しても実施し、データセット B とした。

識別モデルには学習効率と計算コストを考慮して GoogLeNet を採用し、事前学習モデルでは 1000 となっていたネットワークの出力層を 6 クラスに変更し、転移学習を実施した。学習に用いた主要なパラメータを表 3 に示す。

転移学習時の学習曲線の一例を図 5 に示す。上段の縦軸は識別精度、下段は損失関数の値であり、横軸は反復回数（エポック・総学習データ数/バッチサイズ=13200 回）を示している。識別精度では最終的な検証精度が 87.88％に達しているが、さらに向上する余地があると考えられる。損失関数は 0 へ収束する傾向を示しており、学習は安定して進行していることを確認した。

表 2　画像データの枚数

学習用	検証用	評価用
2112	264	264

表 3　転移学習のパラメータ

バッチサイズ	エポック	学習率	最適化アルゴリズム
8	50	1e-5	SGDM

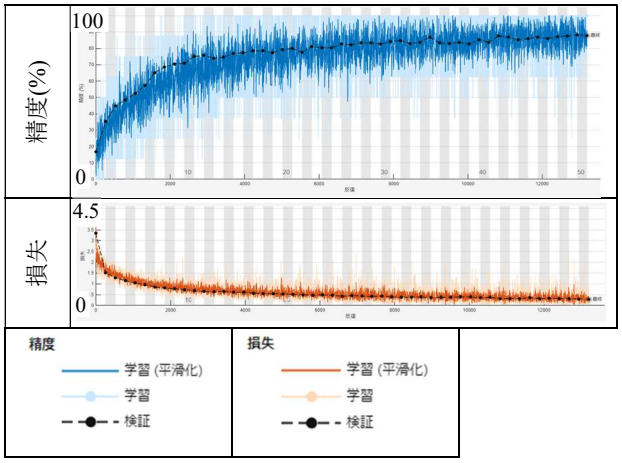

図 5　転移学習による学習曲線（データセット B）

3.3 識別結果

学習済み識別モデルによる評価用データの識別結果の混同行列の一例を図 6 および図 7 に示す。縦軸は予測クラス、横軸は真のクラスを表している。図 6 はデータセット A に対する、図 7 はデータセット B に対する 6 クラス識別を示している。

データセット A の混同行列の識別精度は 90.91％であった。図 6 を見ると、一部の隣接するクラス間で誤識別が発生していることがわかる。一方、noise クラスは真のクラス。なお、同様の実験を 5 回実施した結果、平均識別精度は 90.81%となり、混同行列は図 6 と同様の安定した結果が得られることを確認した。

予測クラス＼真のクラス	0dB	-6dB	-12dB	-18dB	-24dB	noise
0dB	37	0	0	0	0	0
-6dB	7	38	0	0	0	0
-12dB	0	6	44	1	0	0
-18dB	0	0	0	36	2	0
-24dB	0	0	0	7	42	1
noise	0	0	0	0	0	43

図 6　データセット A の混同行列

予測クラス＼真のクラス	0dB	-6dB	-12dB	-18dB	-24dB	noise
0dB	41	4	0	0	0	0
-6dB	3	38	2	0	0	0
-12dB	0	1	33	2	0	1
-18dB	0	1	2	41	2	3
-24dB	0	0	1	0	29	5
noise	0	0	0	1	10	35

図 7　データセット B の混同行列

一方、データセット B の混同行列の識別精度は 84.47%であった。図 7 を見ると、隣接していないクラス間でも一部誤識別が発生している。特に noise に関わる誤識別が多い。同様の実験を 5 回実施した結果、平均識別精度は 84.19%となり、再現性の観点では問題のないことを確認した。

4. AVAS 音検知の検討

本検討では、識別結果を評価する手段として適合率、再現率、F 値を使する。混同行列を True Positive（TP），False Positive（FP），False Negative（FN） に分け、6 クラス分類および 2 クラス分類における各値を SN 比ごとに算出し、評価した。本章では評価手法の目的、手順、評価結果について述べる。

4.1 混同行列による評価

本検討では、6 クラス識別と 6 クラスを雑音と複合音の 2 クラスに見立てた 2 クラス識別を行う。まず、6 クラス識別においては、SN 比ごとの識別精度および誤差の評価を目的とした。次に、2 クラス分類においては、雑音と AVAS 複合音の検出精度の検証を目的とした。

評価の指標として、各クラスの TP，FP，FN を定義し、SN 比ごとの適合率、再現率、F 値を算出した。例として図 8 および図 9 では、6 クラス識別と 2 クラス識別における SN 比 0dB の適合率等を算出する場合の TP，FP，FN をセルの配色によって示している。なお、2 クラス識別では各クラスにおける再現率と適合率の算出に使用する TP の範囲が異なるため、図 9 上段では TP と FP を示し、下段では TP と FP を示す。

本研究においてこれらの指標を採用した理由は次の通りである。適合率は SN 比ごとの識別結果に対する誤検出の少なさを評価するために使用し、再現率は AVAS の有無に対する見落としを評価する目的で使用する。また、それらの指標を総合的に評価するために F 値を使用する。

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{3}$$

予測クラス＼真のクラス	0dB	-6dB	-12dB	-18dB	-24dB	noise
0dB	41	4	0	0	0	0
-6dB	3	38	2	0	0	0
-12dB	0	1	33	2	0	1
-18dB	0	1	2	41	2	3
-24dB	0	0	1	0	29	5
noise	0	0	0	1	10	35

True Positive　False Positive　False Negative

図 8　6 クラス分類における各要素の例

予測クラス＼真のクラス	0dB	-6dB	-12dB	-18dB	-24dB	noise
0dB	41	4	0	0	0	0
-6dB	3	38	2	0	0	0
-12dB	0	1	33	2	0	1
-18dB	0	1	2	41	2	3
-24dB	0	0	1	0	29	5
noise	0	0	0	1	10	35

予測クラス＼真のクラス	0dB	-6dB	-12dB	-18dB	-24dB	noise
0dB	41	4	0	0	0	0
-6dB	3	38	2	0	0	0
-12dB	0	1	33	2	0	1
-18dB	0	1	2	41	2	3
-24dB	0	0	1	0	29	5
noise	0	0	0	1	10	35

True Positive　False Positive　False Negative

図 9　2 クラス分類における各要素の例

4.2 データセット A の評価

データセット A の混同行列から 4.1 の手法によって求めた識別数を表 4 および表 5 に、グラフを図 10 および図 11 に示す。6 クラス分類では、再現率、適合率が SN 比によって上下するものの、F 値は安定した値をとっていることがわかる。2 クラス分類では、-18dB までは再現率、適合率、F 値が 1.00 であり、-24dB では適合率と F 値が少し減少したことがわかる。

4.3 データセット B の評価

データセット B の混同行列から 4.1 の手法によって求めた識別数を表 6 および表 7 に、グラフを図 12 および図 13 に示す。6 クラス分類では、-12dB までの再現率、適合率、F 値がほぼ一定の値を確保していること、-18dB 以降は再現率、適合率が乖離し、F 値の減少していくことがわかる。2 クラス分類では、-12dB までの再現率が 1.00 であり、-18dB 以降は減少傾向が見られる。-6dB までの適合率は 1.00 であり、-12dB 以降は減少傾向が見られる。再現率、適合率の低下に伴い F 値も減少している。

表 4　データセット A の 6 クラス分類

	TP	FN	FP	再現率	適合率	F 値
0dB	37	7	0	0.84	1.00	0.91
-6dB	38	8	7	0.86	0.84	0.85
-12dB	44	0	7	1.00	0.86	0.93
-18dB	36	8	2	0.82	0.95	0.88
-24dB	42	2	8	0.95	0.84	0.89

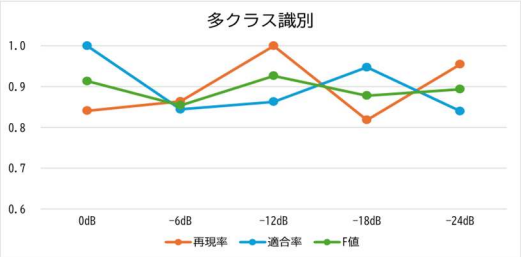

図 10　データセット A の 6 クラス分類

表 5　データセット A の 2 クラス分類

	TP	FN	FP	再現率	適合率	F 値
0dB	44	0	0	1.00	1.00	1.00
-6dB	44	0	0	1.00	1.00	1.00
-12dB	44	0	0	1.00	1.00	1.00
-18dB	44	0	0	1.00	1.00	1.00
-24dB	44	0	1	1.00	0.98	0.99

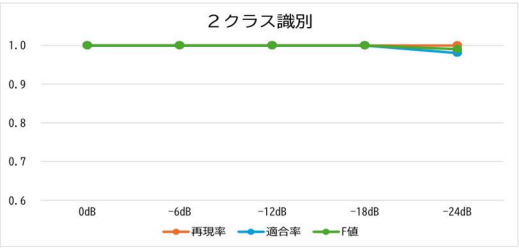

図 11　データセット A の 2 クラス分類

表 6　データセット B の 6 クラス分類

	TP	FN	FP	再現率	適合率	F 値
0dB	41	3	4	0.93	0.91	0.92
-6dB	38	6	5	0.86	0.88	0.87
-12dB	39	5	4	0.89	0.91	0.90
-18dB	41	3	11	0.93	0.78	0.85
-24dB	29	15	6	0.66	0.83	0.73

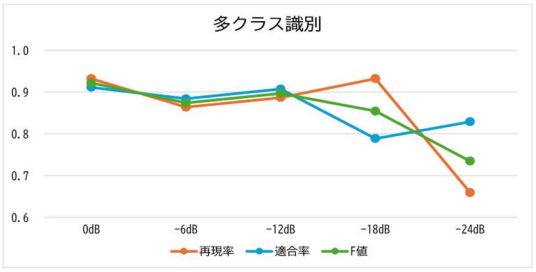

図 12　データセット B の 6 クラス分類

表 7　データセット B の 2 クラス分類

	TP	FN	FP	再現率	適合率	F 値
0dB	44	0	0	1.00	1.00	1.00
-6dB	44	0	0	1.00	1.00	1.00
-12dB	44	0	1	1.00	0.98	0.99
-18dB	43	1	3	0.98	0.94	0.96
-24dB	34	10	5	0.77	0.86	0.81

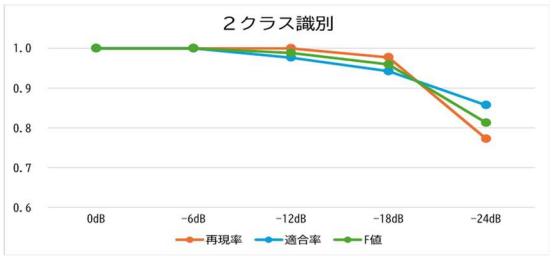

図 13　データセット B の 2 クラス分類

5. まとめと今後の課題

本実験では、深層学習とスペクトログラムを用いることで、SN 比ごと複合音と環境雑音の識別方法について検討した。多クラス分類の評価から、8 割程度の識別ができること、誤識別が発生した際の予測との差は小さいことが分かった。また、2 クラス分類の結果によって、SN 比-18dB まではほぼ安定した識別が行えることが分かった。今後は、モデルのパラメータを最適化して精度の向上を図るとともに、実際の走行データを用いた検証を進めることが課題である。

謝辞

本研究は JSPS 科研費 JP23K11074 の助成を受けたものです。

参考文献

[1] 国土交通省, “「ハイブリッド自動車等の車両接近通報装置」及び「前照灯の自動点灯機能」を義務付けます。”, [2025/6/12 参照]<https://www.mlit.go.jp/report/press/jidosha07_hh_000220.html>

[2] 佐野 将太, 川喜田 佑介, 宮崎 剛, 田中 博, “スペクトログラム画像を用いた転移学習の適用による室内音識別”, 画像電子学会誌, Vol.52, No.2, pp.357-363 (2023)