

スマートフォンを用いた深層学習による 警告音認識システムの改良

矢野和希^{†1} 白石優旗^{†1}

概要：我々は、環境音の中から安全に直結する警告音を確実に認識できるようにするため、深層学習を用いて救急車のサイレンと自転車のベルを識別してユーザに伝達するスマートフォンアプリケーションを開発してきた。本論文では、識別対象にクラクション音と火災警報音を追加するとともに、識別対象以外の音の誤認識への対策も同時に行う。更に、評価実験を行い、5-fold CV 法を用いて識別率およびF 値によりシステムの有効性を検証する。

キーワード：環境音認識, 機械学習, 聴覚障害, 情報保障システム

1. はじめに

日本には、聴覚障害者が約 30 万人存在し、高齢者などの耳がよく聞こえない人を含めると、約 1400 万人存在する [1-2]。それらの耳が不自由な人が、安全・安心に外出できるためには、様々な環境音の中で特に安全・安心な生活に直結する警告音（クラクション、救急車のサイレンなど）を確実に認識できることが求められる。そのため、環境音の中からそれら特定の警告音を識別し、ユーザに伝達するシステムが必要とされている。

一方で、近年、深層ニューラルネットワーク (Deep Neural Network, DNN) という技術が注目されており、認識したい警告音をコンピュータに学習させることで自動的に特徴を取得し、ノイズな環境でもロバストな認識性能を持つと報告されている [3]。それにより対象物の移動や音響環境の変化による音質変化にロバストな高精度の識別が期待される。

そこで、本研究では、深層学習を用いて警告音認識システムを開発する (図 1)。それにより、耳が不自由な人が警告音を確実に認識することができ、安全安心に外出することが可能になる。その際、普及率が平成 27 年末で 78.0% [4] となっており外出時に常に持ち歩くスマートフォンを用いることで、日常的に利用可能なシステムを目指す。

我々は、これまでに、深層学習を用いた警告音認識システムを提案し、その基本的な識別性能について確認し、深層学習を用いて救急車のサイレンと自転車のベルを識別してユーザに伝達するスマートフォンアプリケーションを開発してきた [5-6]。本論文では、識別対象にクラクション音と火災警報音を追加すると共に、識別対象以外の音の誤認識への対策も同時に行う。更に、評価実験を行い、5-fold CV 法を用いて識別率および F 値によりシステムの有効性を検証する。

論文の構成は以下のとおりである。始めに関連研究について述べる。次にシステム概要、識別方法、深層学習によ

る識別器の作成、スマートフォンアプリケーションの動作、スマートフォンを用いた性能評価実験について述べ、最後にまとめと今後の課題について述べる。

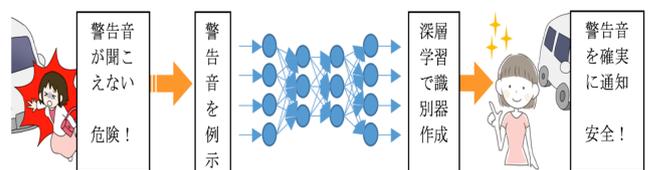


図 1. 提案システム

2. 関連研究

これまでに発表されている警告音認識システムとして、中西らのシステム [7] や岩佐らのシステム [8] などがある。

[7] は、データをサーバに送信し認識するモバイルアプリケーションを開発している。識別手法には、GMM (金剛正規分布モデル) を音響モデルとする音声認識器 Julius を使い、MFCC (メル周波数ケプストラム係数)、 Δ Power を特徴量としているが、平均識別率は 45% 程度であり、識別制度が不十分である。

[8] はパルスニューロンモデルによる識別を行なっている。平均識別率は 95% 程度であるが、自動車に取り付けることを前提としており、歩行時には使用が困難である。また、警告音を発する対象物の移動や周辺環境の変化による警告音の変化への対応が困難といった課題が残されている。

一方で、聴覚障害者のための環境音認識に関する研究に浅井らのシステム [9] がある。[9] は、識別器に SVM (Support Vector Machine) を使い、PLP (Perceptual Linear Predictive) を特徴量としている。平均識別率は 96% 程度であるが、対象音がドアベル、電子レンジの終了音、電話の着信音といった、生活音とされており、警告音認識のような、安全性の観点から非常に高い精度を要求されるタスクの評価はなされていない。

^{†1} 筑波技術大学
Tsukuba University of Technology

3. システム概要

本システムのユーザは、聴覚障害者や聴力の低下した高齢者などであることから、音以外の通知システムが必要となる。本研究では警告音が発生した際に画面に表示する方法を採用する。

提案システムの基本的な流れは以下のとおりである。

- (1) スマートフォンにより環境音を収集
- (2) 警告音識別時にはスマートフォンに通知

識別方法には深層学習を用い、学習データの作成のため、救急車のサイレンや、歩行者や自転車の交通事故防止のためのクラクションやベルなどの通知対象とする音データをあらかじめ収集する。その際、スマートフォンを用いて実環境下において複数の学習データを採取する。また、ノイズな環境音の中で、対象物の移動や音響環境の変化によって音質変化したデータも採取する。ここで、警告音の純音を採取するのではなく、実環境下においてデータを採取する理由は、深層学習の汎化能力を最大限に活用するためである。

様々な環境下で採取した警告音データに対して、データ整理並びに、データスクリーニングを行い、学習用データベースを作成した後、実際に学習を行う。

深層学習ライブラリには Keras[10]を採用する。Keras は Tensorflow[11]のラッパーライブラリであり、Tensorflow 同様にオープンソースでスケラビリティに優れている。また、Linux サーバだけでなく Android, iOS の両スマートフォン OS にも対応しており、開発が容易になる。

4. 識別方法

警告音識別のためには、

- (1) 連続的に環境音を収集
- (2) 閾値以上の音量データを検知した場合、一定時間の音データを記録
- (3) 記録された音データに対して警告種を識別

の3つのステップが必要になる。

また、警告音はその性質上、単調で繰り返される傾向が強いことから、上記の閾値処理により採取された音データに対して短時間フーリエ変換 (Short-Time Fourier Transform, STFT)

$$STFT_{x,\omega}(t,\omega) = \int_{-\infty}^{\infty} x(t)h(\tau-t)e^{-i\omega\tau} d\tau \quad (式1)$$

により、パワースペクトルに変換し、更に log スケールに変換したものを DNN への入力とする。最後に、全ての音データに繰り返し DNN で判断された識別結果に統合処理を適時施すことでリアルタイム識別をする。

5. 深層学習による識別器の作成

先行研究[4]で集めた2種の音データ(救急車のサイレン、自転車のベル)に加え、今回新たに Web ページ[12]よりクラクションの音データ計 18 種類をダウンロードし、また火災警報音を避難訓練の際に録音した。更に、識別対象音以外の音が発生した場合に対応するため、新たに雑音クラスを追加した。今回は、雑音として、足音、車の走行音、声、ドアの開閉音、机を叩く音、ビニール袋の擦れる音の6種を収集した。これらを学習サンプルとして用い、3層 NN, 4層 DNN, 5層 DNN のそれぞれに対して学習、評価を行った。その際、1024[flame]で STFT して得た対数パワースペクトルを NN の入力とし、活性化関数には ReLU を、誤差関数には Softmax 交差エントロピー関数

$$E = -\sum_{c=1}^c \sum_{n=1}^N \{r_{cn} \ln y_{cn}\} \quad (式2)$$

を、学習アルゴリズムには Adam[13]を用いた。

学習・評価用データ 25,000 個 (5,000 個×5 クラス) で CV 法 (5-fold, エポック数 1000) を用い、それぞれの NN に対して、学習、評価を行った。結果を表 1 に示す。

表 1 から分かる通り、いずれの NN についても高い識別結果となったが、今回は最も識別率の高い 5 層 DNN を採用することとした。

表 1. 5-foldCV 法による評価結果

層数	5-foldCV 平均
3層 NN	98.45[%]
4層 DNN	98.67[%]
5層 DNN	99.24[%]

6. 識別アプリケーションの動作

先行研究[5]において、iPhone で識別可能なアプリケーションが開発されている。識別アプリケーションの動作は以下の通りである。

- (1) スマートフォンのマイクロフォンを用い 32bit 単精度浮動小数 (-1.0~1.0) で 1024[flame]毎に集音
- (2) 集音した単精度浮動小数のバッファの絶対値が閾値(0.3)を超えた時に識別処理を開始
- (3) バッファに2の31乗をかけて、バッファの範囲を 32bit 整数型に変えた後、STFT
- (4) 対数パワースペクトルを NN へ入力
- (5) 識別結果を画面表示

実環境においては、対象音は時間的に連続して鳴り続けるため、1回の対象音の発生に対して、識別結果が複数回表示されることになる。したがって、雑音以外の警告音の重要度も考慮し、識別率と可読性を向上させるために以下

のアルゴリズム（統合判定）で最終識別結果を判定する。

- (1) 音を連続評価（1回以上～10回以下）する
 - (2-A) 雑音の以外の特定の音と識別した結果が1回以上ある場合
 - (3-A) 出力の総和を計算する
 - (4-A) 雑音を除いた中から最大のを最終識別結果とする
 - (2-B) 全ての識別結果が雑音であった場合
 - (3-B) 雑音を最終識別結果とする
- 識別結果判定までの全体の動作の流れを図2に示す。

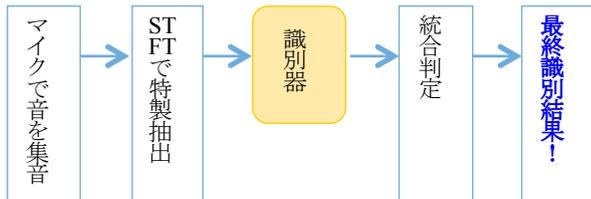


図2. 集音してから最終識別結果を得るまでの流れ

7. 性能評価実験

5節で作成した5層DNNを用いて識別評価を行った。

7.1 室内での評価

最初に、35.7[dB]～43.7[dB]の比較的静かな環境下（屋内）で実験を行った。

今回、救急車のサイレン、火災警報音、クラクション音は実際に鳴らすことが困難なため、別のスマートフォンのスピーカーから過去に録音した音を発生させた。また、雑音としては、足音、車の走行音、声、ドアの開閉音、机を叩く音、ビニール袋の擦れる音の6種について、それぞれ100回（統合判定処理適用後でカウント）ずつ評価した。

6節で述べた統合判定処理適用前の識別結果を表2に、統合判定処理適用後の識別結果を表3に示す。

表2. 静かな環境下における識別結果
 （統合判定処理適用前）

	TP	FP	FN	TN	適合率	再現率	F値
クラクション	558	1	85	3137	0.99	0.86	0.92
自転車のベル	494	0	109	3201	1.00	0.81	0.90
救急車のサイレン	571	2	64	3124	0.99	0.89	0.94
火災警報音	611	1	43	3084	0.99	0.93	0.96
雑音	1167	298	4	2535	0.79	0.99	0.88

表3. 静かな環境下における識別結果
 （統合判定処理適用後）

	TP	FP	FN	TN	適合率	再現率	F値
クラクション	100	0	0	900	1.00	1.00	1.00
自転車のベル	100	0	0	900	1.00	1.00	1.00
救急車のサイレン	100	0	0	900	1.00	1.00	1.00
火災警報音	100	0	0	900	1.00	1.00	1.00
雑音	100	0	0	900	1.00	1.00	1.00

統合判定処理適用後は全ての対象音について精度100%で識別できた。

7.2 実環境下での評価

次に、50.5[dB]～100.3[dB]の騒音の多い実環境下（筑波技術大学天久保キャンパスすぐ側の東大通りの歩道）で実験を行った。

今回、屋外で想定される雑音は6種の中では車の走行音が主であると考え、雑音は走行音のみを100回（統合判定処理適用後でカウント）評価した。その際、それぞれの対象音の音量の最大値も同時に記録した。

6節で述べた統合判定処理適用前の識別結果を表4に、統合判定処理適用後の結果を表5に示す。

表 4. 静かな環境下における識別結果
 (統合判定処理適用前)

	TP	FP	FN	TN	適合率	再現率	F 値	最大音量 [dB]
クラクション	545	0	87	223	1.00	0.86	0.92	98.1
自転車のベル				2				
救急車のサイレン	502	0	113	224	1.00	0.81	0.89	127.7
				9				
救急車のサイレン	572	1	56	233	0.99	0.91	0.95	90.0
				6				
火災警報音	631	1	57	217	0.99	0.91	0.95	93.2
				6				
雑音	298	262	2	256	0.53	0.99	0.69	100.3
				3				

表 5. 静かな環境下における識別結果
 (統合判定処理適用後)

	TP	FP	FN	TN	適合率	再現率	F 値	最大音量 [dB]
クラクション	100	0	0	400	1.00	1.00	1.00	98.1
自転車のベル								
救急車のサイレン	100	0	0	400	1.00	1.00	1.00	127.7
救急車のサイレン	100	0	1	400	1.00	0.99	0.99	90.0
火災警報音	100	0	1	400	1.00	0.99	0.99	93.2
雑音	100	2	0	398	0.99	1.00	0.99	100.3

統合判定処理適用後は平均 F 値 99%以上における精度で実環境において識別できた。

7.3 考察

屋外において雑音の適合率が低くなっている原因について

では、先に述べたように屋外では雑音を車の走行音だけとしているため、それにより雑音の評価回数のみ屋内の時と比べ大きく減ったためである。

雑音の最大値 100.3[dB]は大きなトラックの走行音である。救急車のサイレン、火災警報音とは約 7~10[dB]最大値に差がある。

そこで誤識別が起こった箇所の波形を確認すると、2つの警告音（救急車のサイレン、火災警報音）の最大音量を超える大きな音が発生していることがわかった（図 2）。つまり救急車のサイレンと火災警報音で 1 回ずつ発生した後識別の理由は、警告音を鳴らした直後の大きなトラックの走行音に警告音が埋もれてしまい、雑音として誤識別してしまったものであった。しかし、その前後では正しく警告音と識別しているため、更なる統合処理を施すことで、正しく警告音を伝達可能であった。

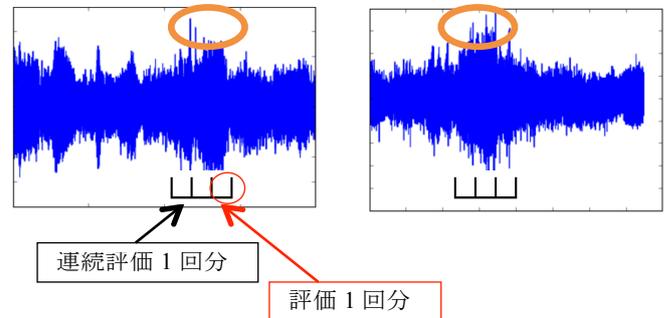


図 2. 誤識別が発生した際の救急車のサイレン（左（連続評価 7 回分））と火災警報音（右（連続評価 8 回分））の波形

7.4 未知のクラクションに対する評価

今回 FFT により特徴抽出、識別を行なっている。特にクラクション音は種類によって音質が変化するため、周波数特性も異なり新しい種類のクラクション音への汎化性能の低さが懸念される。

そこで、最後に、学習データとは異なる新しい種類のクラクション音（20 回*7 種類）に関しても静かな実環境下と騒音の多い実環境下で、6 節で述べた統合判定処理適用前の識別結果（表 6、表 8）と、統合判定処理適用後の結果（表 7、表 9）を確認した。

表 6. 静かな実環境下における
 未知クラクションに対する識別率
 (統合判定処理適用前)

未知クラクション	TP	FN	識別率
クラクション 1	65	11	0.86
クラクション 2	41	14	0.75
クラクション 3	37	9	0.80
クラクション 4	56	24	0.70
クラクション 5	64	7	0.91
クラクション 6	42	22	0.65
クラクション 7	43	36	0.54

表 7. 静かな実環境下における
 未知クラクションに対する識別率
 (統合判定処理適用後)

未知クラクション	TP	FN	識別率
クラクション 1	20	0	1.00
クラクション 2	20	0	1.00
クラクション 3	20	0	1.00
クラクション 4	20	0	1.00
クラクション 5	20	0	1.00
クラクション 6	20	0	1.00
クラクション 7	20	0	1.00

表 8. 騒音の多い実環境下における
 未知クラクションに対する識別率
 (統合判定処理適用前)

未知クラクション	TP	FN	識別率
クラクション 1	62	16	0.79
クラクション 2	43	11	0.80
クラクション 3	42	8	0.84
クラクション 4	55	23	0.71
クラクション 5	67	8	0.89
クラクション 6	36	29	0.55
クラクション 7	50	33	0.60

表 9. 騒音の多い実環境下における
 未学習クラクションに対する識別率
 (統合判定処理適用後)

未学習クラクション	TP	FN	識別率
クラクション 1	20	0	1.00
クラクション 2	20	0	1.00
クラクション 3	20	0	1.00
クラクション 4	20	0	1.00
クラクション 5	20	0	1.00
クラクション 6	20	1	0.95
クラクション 7	20	1	0.95

未知のクラクション音に対しても統合判定処理を施すこと
 によって 95%以上の識別率を得ることができた。

8. まとめと今後の課題

本論文では識別可能な警告音にクラクションと火災警報音を追加し、更に、雑音を第 5 クラスの対象音として学習することで雑音への対策を行なった。結果、騒音の多い実環境下においても、クラクション、自転車のベル、救急車のサイレン、火災警報音全ての場合で 99%以上の識別精度を確認できた。また、未知のクラクション音に対しても 95%以上の識別率が得られた。一方で、警告音と同時に大きな雑音が発生した場合は、警告音が雑音として識別されることがあったものの、更なる統合処理を加えることで正しく警告音を伝達できる可能性が示された。

今後の課題としては、未知のクラクション音の識別率を向上させるため、更なるデータ収集が必要である。その際、クラクションの種類は多種多様であるため、多くの協力者を募ってクラウドソーシングの手法のよりデータ収集を行うことを計画している。また、実用化にあたっては、アプリケーションの改良によるユーザビリティの向上、情報伝達デバイスの活用、集音デバイスの活用などを考慮する必要があると考えている。

謝辞 本研究の一部は、筑波技術大学平成 29 年度学長のリーダーシップによる教育研究等高度化推進事業による助成、並びに JSPS 科研費 JP16K16460 の成果であり、ここに記して謝意を表すものとする。

参考文献

- [1] 平成 25 年版障害者白書 (全体版) 付録障害児・者数の状況, 2013
- [2] 一般社団法人日本歩調工業会, Japan Trak 2015 調査報告, 2015
- [3] N.D.Kane, P.Georgiey, L.Qendro, "DeepEar: Robust Smartphone Audio Sensing in Unconstrained Acoustic Environments using Deep Learning." In Proc. Of the UBICOMP'15, Osaka, Japan, pp. 283-294, 2015
- [4] 総務省, 平成 28 年度版情報通信白書インターネットの普及状況, 2016
- [5] 白石優旗, 深層学習を用いた警告音認識による危険信号通知システムの検討, DEIM Forum 2016 P6-5, 2016
- [6] 畑伸佳, 白石優旗, スマートフォンを用いた深層学習による警告音認識システムの検討, 情報処理学会報告, Vol.2017-AAC-3, No.8, pp.1-4, 2017
- [7] 中西恭介, 津田貴彦, 西村竜一, 河原英紀, 入野俊夫, 松山みのり, 山田順之助, モバイル携帯端末を用いた環境音収集とその認識手法の検討, 情報処理学会研究報告 Vol.2013-MUS-99 No.18, 2013
- [8] 岩佐要, 藤門岳史, クグレマウリシオ, 黒柳奨, 岩田彰, 段野幹男, 宮治正廣, 車載安全運転支援装置のためのパルスニューロンモデルによる音源接近検出および音源種類識別システム, 信学誌, D, 情報・システム, Vol.91, No.4, pp.1130-1141, 2008

- [9] 浅井研哉, 小栗佑介, 志磨村早紀, 北義子, 綱川隆司, 西田昌史, 西村雅史, 聴覚障害者のための実環境下における環境音認識システムに関する検討, 情報処理学会研究報告, Vol.2017-AAC-5, No.11, p1-6, 2017
- [10] Keras Documentation:Keras Documentation (online). available from<<https://keras.io/>>
- [11] Tensorflow Documentation:Tensorflow Documentation (online). available from<<https://Tensorflow.org/>>
- [12] MITSUBA 株式会社ミツバサンコーワ : MITSUBA 株式会社ミツバサンコーワ (オンライン). 入手先 <<http://www.mskw.co.jp/car/car-horn/>>
- [13] Kingma.P.D and Ba, L.: Adam:A Method for Stochastic Optimization, Published as a conference paper at ICLR, Vol.9, No.3, 2015