

# 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

畑伸佳<sup>†1</sup> 白石優旗<sup>†1</sup>

**概要：**我々は、聴覚障害者が自転車のベルや救急車のサイレンなどの警告音を確実に認識できるようにするため、深層学習を用いて環境音の中から特定の警告音を識別してユーザに伝達する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てきた。本論文では、提案システムをスマートフォンに適用可能にすることで、日常的に利用可能なシステムの開発を目指し、評価実験により識別精度を検証する。

**キーワード：** 深層学習, 深層ニューラルネットワーク, 警告音, スマートフォン, パターン認識, 機械学習

## 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てきた [5]。本論文では、提案システムをスマートフォンに適用可能とすることで、日常的に利用可能なシステムの開発を目指し、更に評価実験により識別精度を検証する

論文の構成は以下の通りである。初めに関連研究について述べる。次に、提案システム、識別方法、TensorFlow を用いた学習、スマートフォンアプリケーションの開発、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。

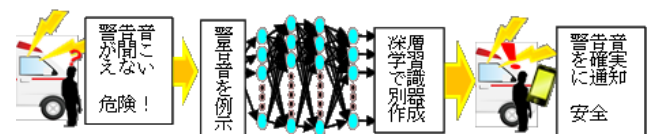


図 1 提案システム

## 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 [6] や岩佐らのシステム [7] がある。

[6] 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 (混合正規分布モデル) を音響モデルとする音声認識器 Julius を用い、MFCC (メル周波数ケプストラム係数)、 $\Delta$  Power を特徴量としているが、平均識別率は 45% 程度であり、識別精度が不十分である。

一方、[7] 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。本研究では、識別手法に深層学習を活用することでこれらの課題の解決を試みる。

## 3. 提案システム

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通りである。

- (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

<sup>†1</sup> 筑波技術大学  
Tsukuba University of Technology

環境下において複数の学習用データを採取する。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 TensorFlow[8]を採用する。なぜなら、TensorFlow はオープンソースでスケーラビリティに優れており、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、開発が容易になると期待されるからである。

#### 4. 識別方法

警告音識別のためには、

- (1) 連続的に環境音を集音
- (2) 閾値以上の音量を検知した場合、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)

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

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 の入力とする。ここで、 $x(t)$  は入力信号、 $t$  は時刻、 $\omega$  は角周波数、 $h(x)$  は窓関数であり、本論文では cosine 窓を用いる。最後に、すべ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. TensorFlow を用いた学習

先行研究[4]で集めた3種の音データ(救急車のサイレン、自転車のベル大、小の2種)を TensorFlow で学習サンプルとして使い、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。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、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[9]を用いる。ここで、 $C$  はクラス数、 $N$  はサンプル数、 $r_{cn}$  は 0 または 1 の教師信号、 $y_{cn}$  は 0~1 の出力信号である。

学習には、学習データ 900 個、評価データ 450 個を用いる。すなわち、本論文では (式2) において、 $C=3$ ,  $N=900$  と設定する。学習の様子を図2に示す。ここで、横軸は学習回数 (epoch)、縦軸は (式2) で計算される誤差である。

評価結果を表1に示す。表1により、すべての NN において識別率 100%を達成していることを確認できる。

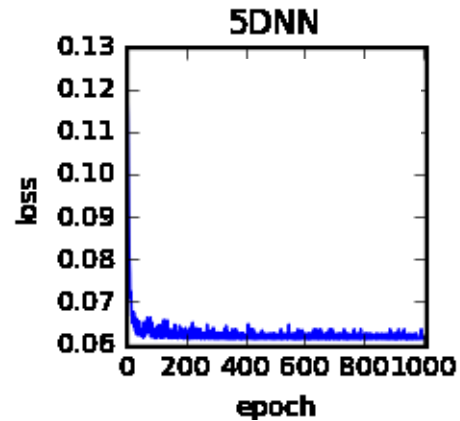


図2 学習曲線

表1 評価結果

	3層 NN	4層 DNN	5層 DNN
識別率	100%	100%	100%

#### 6. スマートフォンアプリケーションの開発

本論文では、プログラミング言語 Swift3 を用いて、iPhone で録音・識別可能なアプリケーションを開発する。

Apple の公開している API BNNS (Basic neural network subroutines) [10]を使い (iOS 10 から使用可能)、スマートフォンで NN を構成する。その NN を使い、警告音認識のためのスマートフォンアプリケーションを作成する。その際、3層 NN, 4層 DNN, 5層 DNN のうちいずれかを選択可能とする。また、識別時に録音するかどうかを選択可能とする。識別アプリケーションの画面を図3に示す。

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ホンを用い 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) を超えたときに識別処理を開始
- (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 に入力
- (5) 識別結果を画面表示

このとき、閾値の設定については、自動車の走行音に反応しないことを基準とする。本論文では、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の車道側に立ち、通常走行時の自動車の走行音に反応しない閾値である 0.3 に

設定した。

本アプリケーションは、識別を行う際に、録音も同時に行うことができる。録音を行うか否かは設定により変更可能である。それにより、識別時の音を後で再現評価することが可能になるとともに、学習サンプルとして再利用することも考えられる。識別時に録音をする録音フォーマットには、非圧縮方式のリニア PCM を採用する。

また、学習サンプルの収集のために録音機能も実装した。その際、録音フォーマットには、変換の際の音の劣化を無くすため、可逆圧縮方式の 1 つである Apple Lossless を採用した。また、ファイル名の変更機能、削除機能も実装し、識別時に録音したファイルも含め管理できるようにしている (図 4)。

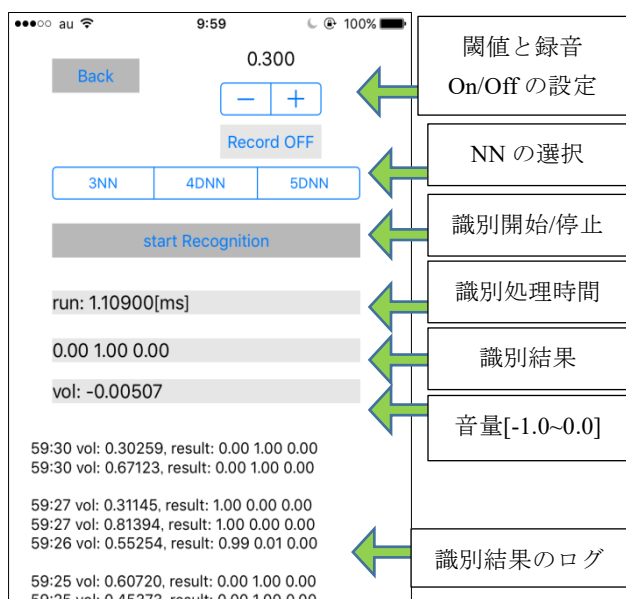


図 3 識別アプリの様子

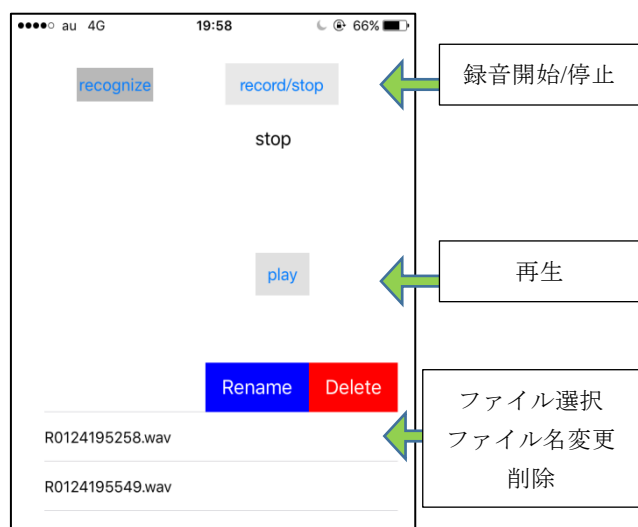


図 4 録音・ファイル管理機能

表 2 屋内での評価実験結果

種類	ベル (大)	ベル (小)	救急車のサイレン音
検知率	100/100 (100%)	100/100 (100%)	100/100 (100%)
識別率	100/100 (100%)	100/100 (100%)	100/100 (100%)

表 3 屋外での評価実験結果：自転車のベル大

距離	25cm	50cm	75cm	100cm	125cm
検知率	100/100 (100%)	100/100 (100%)	94/100 (94%)	84/100 (84%)	71/100 (71%)
識別率	100/100 (100%)	100/100 (100%)	93/94 (99%)	84/84 (100%)	71/71 (100%)

表 4 屋外での評価実験結果：自転車のベル小

距離	25cm	50cm	75cm	100cm	125cm
検知率	100/100 (100%)	95/100 (95%)	93/100 (93%)	89/100 (89%)	92/100 (92%)
識別率	100/100 (100%)	95/95 (100%)	91/93 (98%)	88/89 (98%)	92/92 (100%)

## 7. スマートフォンを用いた性能評価実験

TensorFlow の学習により 3 層 NN で十分な識別率を得られたことから、本論文では 3 層 NN を用いて識別評価を行うこととする。

最初に、比較的静かな環境下 (屋内) で実験を行う。その際、自転車のベル (大, 小), 救急車のサイレン音をそれぞれ 100 回鳴らし、検知率及び識別率について検証する。このとき、自転車のベルはスマートフォンのすぐ側で鳴らすこととした。救急車のサイレン音については、実際に鳴らすことが困難なため PC のスピーカーから過去に録音した音を発生した。

実験結果を表 2 に示す。表 2 から分かる通り、すべての場合において検知率 100% で識別処理が開始された。識別処理が開始された後は、全ての音に対して識別率 100% を確認できた。

次に、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) で実験を行う。実験を行なった当時は、強い風は吹いておらず、歩行者もまばらな状況であった。今回は、自転車のベル (大, 小) をスマートフォンから 25cm, 50cm, 75cm, 100cm, 125cm 離れたところでそれぞれ 100 回ずつ鳴らした。ただし、救急車のサイレン音については、実際に鳴らすことが困難なため実施しな

かった。

自転車のベル大,小についての実験結果をそれぞれ表3,4に示す。これらの表から分かる通り, 検知率は距離が離れるにつれて下がっていく。100cm 離れると, ベル大, 小の平均で87% (193/200) の検知率となった。

表 3,4 から分かる通り, ベル大の平均識別率 99.8% (448/449), ベル小の平均識別率 99.4% (466/469), 全体の平均識別率 99.6% (914/918) であった。なお, 今回実験した範囲では, 識別率と距離に相関は無かった。なお, 誤識別の理由は, ベルを鳴らしたときに大きな自動車の走行音が入ってしまった場合であり, 救急車のサイレンと識別してしまったものであった。

## 8. まとめと今後の課題

本論文では, TensorFlow を使った学習システム, 並びにスマートフォン録音・識別アプリを作成し, 実環境での性能評価実験を行なった。

比較的静かな環境では, 自転車のベル(大・小)と救急車の音に対して, 100%と高い検知率並びに識別率を確認できた。しかし, 騒音のある屋外の実環境においては, 100cm 離れてベルを鳴らしたときの検知率は平均で87%であった。したがって, 今後は閾値の設定を見直す等により検知率100%に近づける。

警告音を検知後の平均識別率は99.6% (914/918) と高い識別率を得ることができた。ただし, 自動車の大きな走行音が同時に入ると誤検知してしまうことがあったため, 自動車の走行音を集めて第4クラスの対象音として学習をし直し, 閾値を下げて自動車の走行音に反応しても通知しない手法についても検討している。また, 警告音の対象を増やすこと, 並びにサンプル音の更なる収集についても行う予定である。また, 今回は識別結果をスマートフォンアプリの画面に提示したが, ウェアラブル端末を用いた振動による通知システムにも拡張予定である。

**謝辞** 本研究の一部は, 筑波技術大学平成 28 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] 内閣府, 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, JapanTrak 2015 調査報告, 2015
- [3] N. D. Lane, P. Georgiev, 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] 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
- [7] 岩佐要, 藤角岳史, クグレマウリシオ, 黒柳奨, 岩田彰, 段野幹男, 宮治正廣, 車載用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及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.
- [8] Google, TensorFlow, <https://www.tensorflow.org/>
- [9] Diederik Kingma, Jimmy Ba, "Adam: A Method for Stochastic Optimization", the 3rd International Conference for Learning Representations, San Diego, arXiv:1412.6980, 2015
- [10] Apple, BNNs, <https://developer.apple.com/reference/accelerate/bnns>