

# 聴覚障害者支援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

浅井 研哉<sup>1,a)</sup> 小栗 佑介<sup>2</sup> 志磨村 早紀<sup>3</sup> 北 義子<sup>4</sup> 綱川 隆司<sup>1</sup> 西田 昌史<sup>1</sup> 西村 雅史<sup>1</sup>

**概要:** 聴覚障害者は音声だけでなく、周囲の環境音が聞こえないという困難を抱えている。室内空間において特定の音を対象として音の到来を障害者に通知する装置が開発されているが、多様な環境音を識別して通知できるものはまだ存在しない。我々は特に装置のコストを抑えるという観点から、スマートフォン1台だけで実現可能な環境音認識手法の検討を行なっている。対象音が多様であるため、ここでは使用者自身が対象音を事前登録する方法を取るが、学習時と認識時の音響ミスマッチによる認識性能の劣化が生じやすいという課題があった。今回、そのようなミスマッチに対処するため、PLP 特徴量の導入を検討した。本報告では3種類の実環境下で生活環境音をスマートフォンを用いて録音し、認識実験を行った結果について報告する。

**キーワード:** 環境音認識, 音情報処理, 機械学習, 聴覚障害者, 音響イベント検出

## A study on Environmental Sound Recognition System in a Real Environment for Hearing Impaired

KENYA ASAI<sup>1,a)</sup> YUSUKE OGURI<sup>2</sup> SAKI SHIMAMURA<sup>3</sup> YOSHIKO KITA<sup>4</sup> TAKASHI TSUNAKAWA<sup>1</sup>  
MASAFUMI NISHIDA<sup>1</sup> MASAFUMI NISHIMURA<sup>1</sup>

**Abstract:** Hearing impaired people have difficulty not only to hear people's voice but also to notice environmental sounds. Some Devices that notify the occurrence of one specific sound to disabled people in indoor spaces have already been developed. But there is no device that can identify and notify various arbitrary environmental sounds. We are studying environmental sound recognition method that works with one smart phone to reduce device cost. Since the target environmental sound is varied, user needs to pre-register the target sound, there was a problem that the recognition performance tends to deteriorate because of the acoustic mismatch with train data and test data. To cope with such a mismatch, we considered the introduction of PLP feature. In this paper, we report the results of recognition experiment, using the living environment sound data that collected by smartphone, conducted under three kinds of real environments.

**Keywords:** Environmental Sound Recognition, Sound Information Processing, Machine Learning, Hearing Impaired, Audio Event Detection (AED)

<sup>1</sup> 静岡大学大学院  
Graduate School of Shizuoka University  
<sup>2</sup> 静岡大学  
Shizuoka University  
<sup>3</sup> 早稲田大学 障がい学生支援室  
Waseda University Office for Students with Disabilities  
<sup>4</sup> 国立障害者リハビリテーションセンター学院  
National Rehabilitation Center for Persons with Disabilities  
a) asai.kenya.17@shizuoka.ac.jp

## 1. はじめに

聴覚障害者は会話などの音声に加えて、生活を送る上で重要な生活環境音（着信音、ヤカン警笛音等）が聞こえないという困難を抱えている<sup>[1]</sup>。そのため、聴覚障害者を支援する聴導犬という介助犬が存在する。しかし、聴導犬の育成には長い期間と多くの費用が必要であることと、聴導

犬を利用するための利用条件に障害者程度等級2級（聴覚障害では最高等級）が要されることから、現在は聴導犬の普及が困難となっている。数値としては、潜在的な聴導犬の希望者数10,000人に対して、2017年9月1日現在の聴導犬の実働数は71頭であり、需要に対して供給が追いついていない<sup>[2][3]</sup>。また、聴導犬のようなサポートは受けたいが、飼育の手間や家屋の設備等の問題のため、そもそも自宅で聴導犬を利用することが困難な状況である場合も多い。そのような聴覚障害者にとって、手軽に誰でも利用することが可能なサポートシステムは利便性が高いと言える。

そこで、居住空間内における生活環境音をシステムにより自動認識し、それが何の音であったのかを聴覚障害者に通知するというシステムが考えられる。また、環境音認識システムによって聴覚障害者自身の障害認識を可能にし、QOL(Quality of life)の向上に繋がることも期待出来る。

環境音認識に関する研究は以前から多く行われているが、必ずしも聴覚障害者支援を意図したものではなかった。また、実際にシステムを利用することを想定し、実環境下で起こり得る問題について検討している研究は少ない。よって、本報告では環境音認識システムの実現に向け、実環境下で発生する状況が認識結果に与える影響についての確認を行うとともに、実使用環境下での性能改善の検討を行った。

本報告は次のように構成される。2章では、環境音認識の関連研究についてまとめる。3章で提案手法について述べた後、4章、5章で実験概要および実験結果を示す。最後に6章で結論を述べる。

## 2. 関連研究

環境音認識に関する研究は数多く行われている。それらの研究を大別すると、深層学習を用いるもの<sup>[4-14]</sup>と、深層学習を用いないもの<sup>[15-26]</sup>に分けることができ、研究の幅は広い。なお、深層学習を用いない環境音認識の論文では、特徴量としてMel-frequency cepstral coefficients (MFCC)を扱っている研究が多く見られる<sup>[18-26]</sup>。また、3章でも後述するが、文献<sup>[24-26]</sup>の中では、Perceptual Linear Prediction (PLP)も特徴量として扱われており、本報告の提案法と共通する。しかし、本報告の3章で述べているような少量のデータによる登録方法を採用している研究はいずれも存在しない。

関連研究の中でも、DCASE 2016(Detection and Classification of Acoustic Scenes and Events 2016)で発表された文献<sup>[23]</sup>では、モバイルデバイスを用いた環境音認識システムについての検討を行っており、本報告の最終目標と共通する部分が多く存在する。Pillosらの方法<sup>[23]</sup>は、Android OSを利用したモバイルデバイス上に環境音認識システムを実装しており、DCASEで配布されている10クラスの環境音データセットによるリアルタイムな認識実験が行われている。特徴量にはZero Crossing

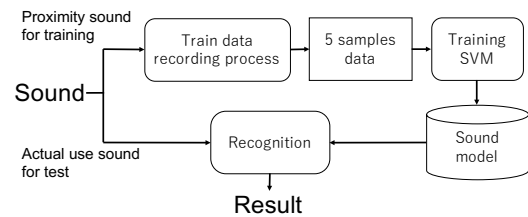


図1 システム構成図

Fig. 1 System configuration

Rate, Spectral Centroid, Spectral Flatness, MFCCが利用されている。また、識別器にはMulti-layer Perceptron (MLP), Support Vector Machine (SVM), RandomForest, BayesNet, NaiveBayesが検討されており、MLPが最も良い性能であったと報告されている。しかし、研究目的に聴覚障害者支援が含まれているにも関わらず、用いたデータセットがヘリコプターのプロペラ音、チェーンソーの稼働音、海の波、鳥の鳴き声といったものが多く、聴覚障害者にとってあまり実用的なデータではないと論文内で言及されている。

そこで本報告では、聴覚障害者が実際にシステムを利用する場合を想定したデータセットでの環境音認識実験について、また、実環境下においてMFCCよりも頑健な特徴量について、これら2つの観点から議論を行う。

## 3. 提案手法

本報告における提案システムの構成図を図1に示す。提案するシステムでは、使用者が事前にスマートフォンのマイクで録音した数サンプルのデータだけで、対象音を登録できるというシステムを想定している。また、文献<sup>[23]</sup>でも述べられているように、常に音データをサーバとやり取りするのは多くのインターネット帯域幅を要することになり、現実的ではない。本報告が想定する最終的なシステムの形式では、特徴抽出や音響モデルの学習といったシステム自体も録音を行なったデバイス上で動作することが望ましい。その場合、深層学習のように膨大な学習データが必要となるシステムは好ましくないため、今回は深層学習を用いない手法に焦点を当てて研究を行なった。

### 3.1 音響特徴量について

提案法の音響特徴量にはPerceptual Linear Predictive(PLP)を次数24として用いる<sup>[27]</sup>。2章でも述べたように、環境音認識の関連研究では音響特徴量としてMFCCが多く用いられているが、音声認識の分野においてはMFCCよりもPLPの方が音響的なミスマッチに頑健であるとされている<sup>[28]</sup>。

実際のシステム利用時には、学習時と実使用時の収録環境が異なることが予想される。すなわち、学習データとテストデータのデータ間の音響ミスマッチが大きくなること

表 1 選定した対象環境音

Table 1 Selected environmental sound

Environmental sound	Class name
笛吹きケトルの警笛音	boiling
来客時のドアベル	doorbell
電子レンジの終了音	microwave
電話の着信音	ringtone
目覚まし時計 (ベル式)	alarm

が懸念されるため、それに耐えうる特徴量として PLP の利用を試みた。PLP 抽出時のフレーム長は 60 ms, フレームシフト幅は 10 ms とした。

### 3.2 識別器について

提案するシステムでは学習データが少量になることが予想される。そこで識別器には、少ない学習データで動作が可能な Support Vector Machine (SVM) を用いて、フレーム単位での学習および識別を行った。カーネルには RBF カーネル (Gaussian カーネル) を利用した。また、様々な環境に汎用的に対応することが求められるため、パラメータチューニングは行っていない。

### 3.3 統合化について

SVM によるフレーム単位の識別結果に対して 2 段階のスムージングを行うことで平滑化を行う。

1 段階目では、あるフレーム  $i$  を中心とした  $i-n \sim i+n$  フレームの区間に対してフレーム単位識別結果の最頻値を求め、 $i$  フレーム目の結果とする。1 段階目においては、スムージングする際のシフト幅は 1 フレームである。

2 段階目では、1 段階目で平滑化した結果に対し、ある  $i$  フレームを含めた  $i-n/2 \sim i+n/2$  で最頻値を求め、 $i-n/2 \sim i+n/2$  フレームの結果とすることで平滑化を行う。2 段階目においては、スムージングする際のシフト幅は  $n/2$  フレームである。

1 段階目のスムージングでは主に突発的な非定常雑音への対処を、2 段階目のスムージングでは 1 段階目で完全に対処しきれなかった誤識別への対処を行なっている。なお、今回  $n$  の値は 50 として処理を実行した。

## 4. 実験概要

### 4.1 対象環境音の選定

居住空間内における環境音は多様であるので、今回は認識対象とする環境音の選定を行なった。表 1 に認識対象とした環境音と、そのクラス名を示す。

### 4.2 録音環境

聴覚障害者が実際にシステムを利用することを想定し、本報告では 2 種類のマンションと 1 種類のオフィス、計 3 種類の実環境下にて音データの録音を行なった。2 種類の

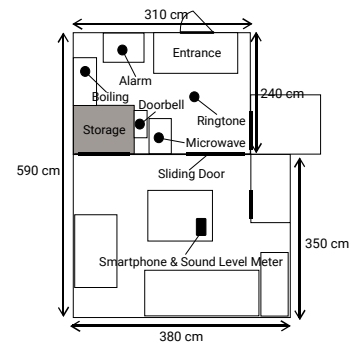


図 2 実験環境 (a): 各音源と測定位置, ドアは閉め切った状態で収録

Fig. 2 Experimental environment (a): Each sound source and recording position, the door is closed

マンションである実験環境 (a), (b) のレイアウト図をそれぞれ図 2, 図 3 に、オフィスである実験環境 (c) のレイアウト図を図 4 に示す。図中にある黒丸の位置に各音源を発生させる家電や機器を配置した。学習データは音源の近くで録音し、テストデータは図中の定位置にスマートフォンを固定して録音した。

各実験環境の備考について説明する。実験環境 (a), (b) については、換気扇を動作させたまま、部屋間のドアや引き戸は閉め切った状態で録音を行なった。そのため、テストデータの録音時は、全ての認識対象音とスマートフォンのマイクの間には遮蔽物が存在することとなり、間接音での録音となる。実験環境 (c) についても、換気扇・空調を動作させたまま録音を行なった。また、実験時に在室していた 14 名には普段通りに活動してもらった。そのため、実験環境 (c) においては話し声や物音 (キーボードのタイプ音, ドアの開閉音, 椅子の転む音等) といった雑音が多い環境下になっている点が、実験環境 (a), (b) との差異である。

また、どの実験環境のテストデータも外部からの雑音 (車・バイクの走行音, 救急車のサイレン, 工事の作業音等) が混入している区間が存在している。

### 4.3 実験データの内訳

識別器の学習データには、音源から約 20 cm の至近距離で録音した音データを各クラス 5 サンプル用いた。(boiling クラスのみ 1 サンプル) テストデータには、4.2 で述べたように図 2, 図 3, 図 4 の図中の位置にスマートフォンを固定して録音したデータを使用した。テストデータは、3 分ごとに各クラスの対象音を 1 サンプルずつ発生させ、それを 1 時間継続して行うという録音方法を各実験環境で行なった。そのため各実験環境の 1 時間のテストデータ中には、5 種類の各対象音が 20 サンプルずつ (計 100 サンプル) と、対象音以外の音 (定常雑音, 非定常雑音) が含まれていることになる。

音データは Linear PCM 形式, モノラル, サンプリング

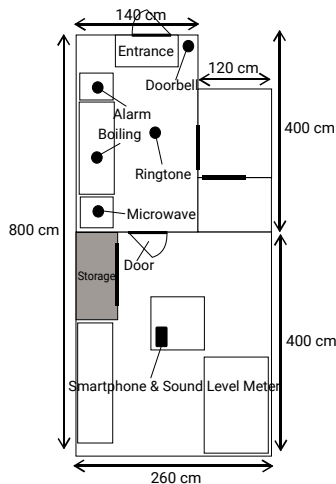


図 3 実験環境 (b): 各音源と測定位置, ドアは閉め切った状態で収録

Fig. 3 Experimental environment (b): Each sound source and recording position, the door is closed

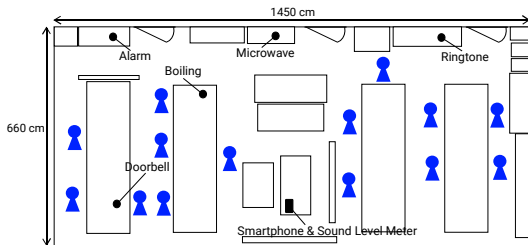


図 4 実験環境 (c): 各音源と測定位置, 及び実験中に在室していた人が主に居た位置を示す

Fig. 4 Experimental environment (c): Each sound source, recording position and where the person who stayed in the experiment was mainly stayed position

表 2 実験に用いた音データの時間長 (単位: 秒)

Table 2 Time length of sound data (unit: sec)

Data		Train data			Test data		
Experimental environment		(a)	(b)	(c)	(a)	(b)	(c)
Sound class	boiling	10.9	12.3	11.3	189.3	175.3	212.8
	doorbell	9.8	9.8	9.8	40.0	30.0	40.0
	microwave	14.8	3.2	4.8	60.0	54.0	16.0
	ringtone	3.2	2.7	3.3	14.0	14.0	14.0
	alarm	12.2	12.3	12.3	49.0	50.0	50.0
	other(noise)	10.0	10.0	10.0	3120.4	3119.2	3183.9

周波数 48,000 Hz, 量子化 16 bit で収録した。また, 雑音クラス (other クラス) の学習には, 各実験環境におけるテストデータ中の 10 秒の定常雑音をそれぞれ用いた。各実験環境の実験データの時間長を表 2 に示す。

#### 4.4 実験データの音圧レベル

各認識対象音および雑音に対して, 音圧レベルを測定した。測定には騒音計を用いており, フラットな周波数特性である C 特性を利用した。各実験環境において 1 時間計測した時の平均音圧レベルを, 学習データについては図 5 に, テストデータについては図 6 に示す。なお, Stationary

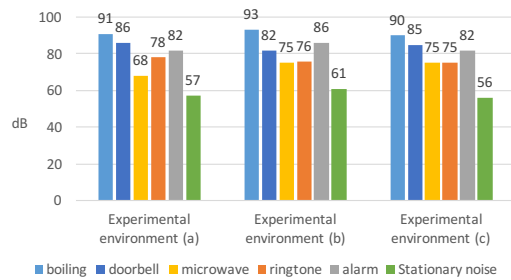


図 5 学習データの音圧レベル

Fig. 5 Sound level of training data

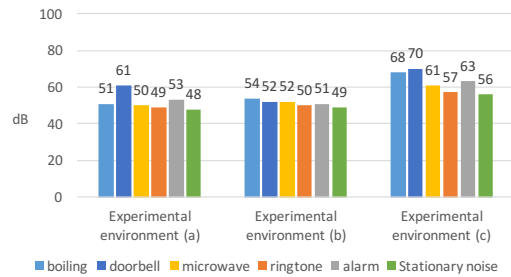


図 6 テストデータの音圧レベル

Fig. 6 Sound level of test data

noise は各実験環境の定常雑音を表す。

## 5. 実験結果

本節では, 提案手法を用いて環境音認識実験を行なった結果について議論する。評価指標には, Precision と Recall の調和平均である F-measure を用いた。

### 5.1 フレーム単位での性能評価

各実験環境における, フレーム単位での識別を行なった結果を図 7 に示す。図中の棒グラフの数値はそれぞれ, 実験環境毎に各クラス (boiling, doorbell, microwave, ringtone, alarm, other) の F-measure を平均した値となっている。どの実験環境においても, MFCC よりも PLP の方が高い精度となっていることが読み取れる。特に, マンションである実験環境 (a), (b) は, オフィスである実験環境 (c) よりも MFCC と PLP の性能差が大きくなっていることが分かる。これは音源の配置距離の関係だけでなく, 4.2 章でも述べたように, 実験環境 (a), (b) のテストデータは全て扉越しに録音された間接音であることが影響していると考えられる。また, 図 5, 図 6 にも示したように実験環境 (a), (b) は, 音圧レベルの観点から見ても学習データとテストデータにおける音響ミスマッチが比較的大きいと言える。

### 5.2 イベント単位での性能評価

実用的な性能評価のため, フレーム単位の性能評価だけでなく, イベント単位での評価も併せて行なった。各実

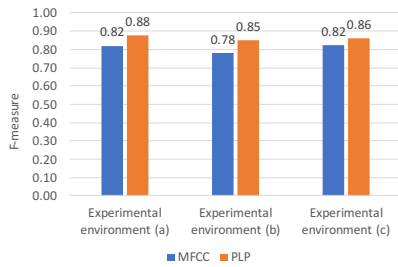


図 7 フレーム単位による認識精度の結果

Fig. 7 Results of recognition accuracy by sound frame

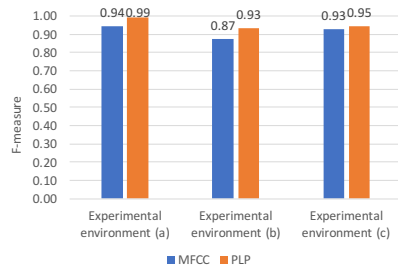


図 8 イベント単位による認識精度の結果

Fig. 8 Results of recognition accuracy by sound event

験環境における、イベント単位による識別結果を図 8 に示す。図中の棒グラフの数値はそれぞれ、実験環境毎に各クラス (boiling, doorbell, microwave, ringtone, alarm) の F-measure を平均した値となっている。

なお本報告における、イベント単位識別時の正解と不正解の基準について述べる。

**正解の基準：** 認識対象音が発生した区間 (正解区間) とその前後 0.5 秒を含めた範囲内に、正解区間の半分以上の長さで正しい結果が推定できている場合に正解としてカウント。

**不正解の基準：** 正解区間や雑音区間を問わず、誤った結果が推定されている場合にカウント。

図 8 から、5.1 章の結果と同様、全ての実験環境において MFCC より PLP の方が高い性能となっていることが分かる。また、実験環境 (c) よりも、実験環境 (a), (b) の方が、MFCC と PLP の性能差が大きくなっていることもフレーム単位識別時と同様である。

各実験環境における、イベント単位での識別を行なった場合の誤識別の内訳を表 3 に示す。各クラス名の右にある括弧内の数字は、そのクラスにおける誤識別の回数を表す。誤識別の主な原因としては、救急車のサイレン音、ドアの開閉音、人の笑い声といった非定常雑音の影響が挙げられる。

MFCC と比べると PLP は全体的な誤識別の回数が改善されている。また、発生する誤識別のパターンにも差異があった。例えば実験環境 (c) においては MFCC を用いた場合に、人の笑い声が doorbell や microwave に誤識別されていたが、PLP ではそのような誤識別のパターンは見られ

表 3 イベント単位識別時に誤識別した音の内訳

Table 3 Details of identification error by sound event

Feature kind	Experimental environment	Sound class of identification error
MFCC	(a)	microwave(6), ringtone(5), alarm(1)
	(b)	ringtone(1), alarm(23)
	(c)	boiling(6), doorbell(1), microwave(8)
PLP	(a)	microwave(1), alarm(1)
	(b)	microwave(1), alarm(12)
	(c)	boiling(7), microwave(2), alarm(1)

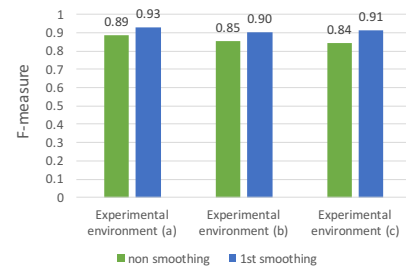


図 9 フレーム単位識別時のスムージングの有無による性能比較

Fig. 9 Performance with and without smoothing in sound frame recognition

なかった。

### 5.3 統合化による認識結果への影響

ここでは統合化の有無や、スムージングの段階別による性能の比較を行う。なお、本報告における実験では MFCC よりも PLP の方が高い性能であったため、本節での結果は特徴量に PLP を用いた場合のみを掲載する。

フレーム単位識別における、統合化の有無による比較を図 9 に示す。図中の棒グラフの数値はそれぞれ、実験環境毎に各クラス (boiling, doorbell, microwave, ringtone, alarm, other) の F-measure を平均した値となっている。図 9 の結果を見ると、1 段階目のスムージング (1st smoothing) の適用で、全実験環境において 9 割を超える性能で識別が行えていることが分かる。しかし、1 段階目のスムージングのみではイベント単位識別時における細かな誤識別に対処できないパターンが見られた。その改善を目的とし、3.3 章でも述べたように 2 段階目のスムージング (2nd smoothing) を行なった。

イベント単位識別における、スムージングの段階別 (1 段階目, 2 段階目) による比較を図 10 に示す。図中の棒グラフの数値はそれぞれ、実験環境毎に各クラス (boiling, doorbell, microwave, ringtone, alarm) の F-measure を平均した値となっている。実験環境 (b), (c) において、2 段階スムージングによる性能改善が見られる。

## 6. おわりに

本報告では、3 種類の異なる実験環境下において環境音認識システムの精度を確認し、システムの認識性能を高める

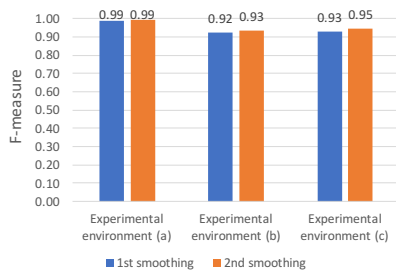


図 10 イベント単位識別時のスムージングの段階による性能比較  
 Fig. 10 Performance with each smoothing stage in sound event recognition

ための検討を行なった。実環境下においては、関連研究で多く用いられている MFCC よりも PLP の方が認識性能が高くなる傾向にあった。また、本報告の提案手法を用いることで、少量の学習データで、かつスマートフォン等のデバイスが 1 台あれば、平均 F 値 0.9 以上の精度で実環境下における環境音を認識できる可能性が示唆された。

今後は、様々な利用環境に対応することを目的として、システム利用者が認識感度を調節できるような仕組みについての検討を行なっていきたい。

#### 参考文献

[1] 勝谷紀子, “難聴者が日常生活で経験するストレスとは”, 日本心理学会大会発表論文集, 76th, pp.386, 2012.

[2] 厚生労働省, 補助犬の実働頭数入手先 (<http://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000165273.html>)

[3] 公益財団法人, 日本補助犬協会, 補助犬アカデミー Q&A 入手先 (<http://www.hojyoken.or.jp/qa/>)

[4] 鳥羽隼司, 原直, 阿部匡伸, “スマートフォンで収録した環境音データベースを用いた CNN による環境音分類”, 日本音響学会講演論文集 (春), 2-P-2, pp.139-142, 2017.

[5] O. Gencoglu, T. Virtanen, H. Huttunen, “Recognition of acoustic events using deep neural networks”, European Signal Processing Conference (EUSIPCO), 2014.

[6] E. Miquel, F. Masakiyo, K. Keisuke, N. Tomohiro, “Exploiting spectro-temporal locality in deep learning based acoustic event detection”, EURASIP Journal on Audio Speech and Music Processing, 2015.

[7] N. D. Lane, P. Georgiev, L. Qendro, “DeepEar Robust Smartphone Audio Sensing in Unconstrained Acoustic Environments using Deep Learning”, UbiComp, pp. 283-294, 2015.

[8] K. J. Piczak, “Environmental sound classification with convolutional neural networks”, Machine Learning for Signal Processing (MLSP), 2015.

[9] 川西 誠司, サクティ サクリアニ, 吉野 幸一郎, ニュービッググラム, 中村 哲, “Deep Neural Network を用いた音声と環境音のマルチタスク学習”, 日本音響学会講演論文集 (春), 3-P-6, pp.163-164, 2016.

[10] 美島咲子, 水野智之, 若林佑幸, 福森隆寛, 中山雅人, 西浦敬信, “複数フレームのフィルタバンクを用いた深層ニューラルネットワークによる室内環境音識別の性能評価”, 日本音響学会講演論文集 (秋), 2-Q-1, pp.39-40, 2016.

[11] 篠崎隆宏, “聴覚情報保障のための生活下トリガー音認識システムの研究”, 電気通信普及財団研究調査報告書,

No.28, pp. 484-492, 2013.

[12] 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, pp. 5-6, 2016.

[13] 美島 咲子, 若林 佑幸, 福森 隆寛, 中山 雅人, 西浦 敬信, “時間波形を用いた深層ニューラルネットワークによる室内環境音識別の検討”, 日本音響学会講演論文集 (春), 2-5-4, pp.43-44, 2017.

[14] 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”, 研究報告アクセシビリティ (AAC-3), No. 8, pp. 1-4, 2017.

[15] 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”, 電子情報通信学会論文誌, Vol. J91-D, No. 4, pp. 1130-1141, 2008.

[16] 根岸 佑也, 河川 信夫, “高度な実世界イベント認識を手軽に利用可能にする Instant Learning Sound Sensor の提案”, 情報処理学会論文誌, Vol. 50, No. 4, pp. 1272-1286, 2009.

[17] 猿舘 朝, 布川 博士, 伊藤 憲三, “携帯端末を利用した難聴者向け生活音サポートシステム”, 日本感性工学会論文誌, Vol. 15, No. 1, pp. 97-105, 2016.

[18] E. Miquel, F. Masakiyo, N. Tomohiro, “Detection and classification of acoustic events using multiple resolution spectrogram patch models”, 日本音響学会講演論文集 (秋), 3-8-4, pp.1529-1530, 2014.

[19] A. Mesaros, T. Heittola, A. Eronen, T. Virtanen, “Acoustic event detection in real life recordings”, 18th European Signal Processing Conference (EUSIPCO), pp. 1267-1271, 2010.

[20] Z. Zhao, S. Zhang, Z. Xu, K. Bellisario, N. Dai, H. Omrani, B. C. Pijanowski, “Automated bird acoustic event detection and robust species classification”, Ecological Informatics, 2017.

[21] C. V. Cotton, D. P. W. Ellis, “Spectral vs. spectro-temporal features for acoustic event detection”, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2011.

[22] 佐々木洋子, 吉井和佳, 加賀美聡, “無限混合ガウスモデルを用いた未知クラスに対応可能な実環境音分類法”, 人工知能学会 AI チャレンジ研究会, 36th, pp. 7, 2012.

[23] A. Pillos, K. Alghamidi, N. Alzamel, V. Pavlov, S. Machanavajhala, “A REAL-TIME ENVIRONMENTAL SOUND RECOGNITION SYSTEM FOR THE ANDROID OS”, Detection and Classification of Acoustic Scenes and Events (DCASE), 2016.

[24] I. Trancoso, J. Portelo, M. Bugalho, J. Neto, A. Serralheiro, “Training audio events detectors with a sound effects corpus”, Proc. Interspeech, 2008.

[25] J. Rouas, J. Louradour, S. Ambellouis, “Audio Events Detection in Public Transport Vehicle”, Intelligent Transportation Systems Conference, 2006.

[26] M. Bugalho, J. Portelo, I. Trancoso, T. Pellegrini, A. Abad, “Detecting Audio Events for Semantic Video Search”, Proc. Interspeech, pp. 1151-1154, 2009.

[27] H. Hermansky, “Perceptual linear predictive (PLP) analysis of speech”, The Journal of the Acoustical Society of America, Vol. 87, No. 4, pp. 1738-1752, 1990.

[28] 鈴木 雅之, “背景雑音と話者の違いに頑健な音声認識”, 博士学位論文, 東京大学, 2013.