

聴覚障害者支援のための環境音可視化システムの開発

浅井 研哉^{1,a)} 綱川 隆司¹ 西田 昌史¹ 西村 雅史¹

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。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や、周囲の騒音レベルに応じて振動するデバイスなどが開発されているが、実際に聴覚障害者が利用しているものはあまり多くない。本報告では、周囲の音を自動認識して必要な情報のみを聴覚障害者に通知する環境音可視化システムを提案し、その実現のための基礎的な検討を行った。これまで我々は、特に装置のコストを抑えるという観点から、スマートフォンやタブレットといった端末の内蔵マイクを用いた環境音認識についての研究を行ってきた。システムの認識対象音をカスタマイズできる方式を採用し、聴覚障害者が必要とする音だけを認識することで、マイクロフォンアレイなどの特別な機器を必要とせずに実用性の高いシステムの実現を図る。今回、実際に実時間動作可能な環境音可視化システムをタブレット端末上に開発した。実験では、健聴者を対象とした主観評価による有効な情報提示方法の調査と、システムの認識性能の評価を実施した。

キーワード: 環境音認識, 音情報処理, 機械学習, 障害者支援, 音イベント検出

Development of Environmental Sound Visualization System for Hearing Impaired

KENYA ASAI^{1,a)} TAKASHI TSUNAKAWA¹ MASAFUMI NISHIDA¹ MASAFUMI NISHIMURA¹

Abstract: Hearing impaired people have difficulty not only to hear human voices but also to notice environmental sounds. Some Devices that notify the occurrence of one specific sound to disabled people and vibrates in response to the magnitude of the surrounding sound have already been developed. However, there are not many devices actually used by hearing impaired. In this report, we propose an environmental sound visualization system. Which automatically recognizes environmental sounds and notifies only the necessary information to hearing impaired. In this system, it is also possible for hearing impaired to select target sounds which they want to know. Actually, we developed an environmental sound visualization system that can work in real time on a tablet without using a special device such as microphone array. In experiments, we conducted a survey of effective visualization method and evaluated recognition performance of the system.

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

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環境音が聞こえないという困難を抱えている^[1]。そこで、身の回りの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通知するという環境音可視化システムが検討されている。環境音可視化システムは、例えば補聴器が音声以外の音を抑制しているように、認識を通じて余計な情報を消し、必要な情報のみを伝えることを目的としている。

本研究では、機械学習の技術を用いて環境音認識を行なう。環境音認識についてはこれまで多くの研究が行われてきたが、聴覚障害者支援を意図したものは多くはなかった。

¹ 静岡大学大学院総合科学技術研究科情報学専攻
Shizuoka University, Graduate School of Integrated Science and Technology, Department of Informatics

^{a)} asai.kenya.17@shizuoka.ac.jp

加えて、実際に聴覚障害者がシステムを利用することを想定し、そのシステムの有用性や実用性について検討している研究は少ない。そこで、我々は一般的な環境音認識を目指すのではなく、聴覚障害者が日頃立ち寄り場所が必要となる環境音を、事前にシステムに登録するという方法を採用した。利用者によるカスタマイズが可能なシステムにすることで、一般的な環境音認識システムでは実現できない高い認識精度を実現するシステムを目指している。本報告では、そのような仕組みを取り入れた環境音可視化システムの基礎的検討として、タブレット上で実時間動作可能なシステムを開発し、情報提示方法の検討および認識性能の評価を行なった。

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や、自身が行なってきた研究について紹介する。3章で提案手法について述べた後、4章で実験の概要および結果を示す。最後に5章で結論と今後の展望を述べる。

2. 関連研究

環境音認識については、数多くの研究が行われている。それらの研究を大別すると、深層学習を用いるもの^[2-12]と、深層学習を用いないもの^[13-24]などに分けることができ、研究の幅は広い。

この中でも、DCASE 2016 (Detection and Classification of Acoustic Scenes and Events 2016) では、モバイルデバイスを用いた環境音認識システムが数多く報告されている。Pillos ら^[21]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、DCASE で配布されている 10 クラスの環境音データセットを用いてリアルタイム認識実験を行なった。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はチェーンソーの稼働音、ヘリコプターのプロペラ音、鳥の鳴き声、海の潮騒といったものであり、聴覚障害者にとってあまり実用的なデータではないと、この論文内でも言及されている。

一方、我々は特に装置のコストを抑えるという観点から、スマートフォンやタブレットといった端末のアプリケーションとして動作し、認識対象音を簡便に登録することができる認識システムについて検討を行ってきた^[25,26]。ここでは、実環境下であるマンションやオフィスを対象とし、室内空間における環境音認識について取り組んだ。実験では、スマートフォンのマイクを用いて、ドアベルや着信音といった認識対象の近接音を事前に収録し、学習モデルを構築した。その後、スマートフォンを隣室や認識対象音から離れた場所に設置し、マイクと音源の間に遮蔽物があるような状態で環境音認識実験を行い、フレーム単位識別で平均 F-measure 0.9 以上の認識性能を確認した。このシステムでは、対象音の登録時のみ健聴者のサポートが必要となるが、対象を限定することで、マイクロフォンアレイな

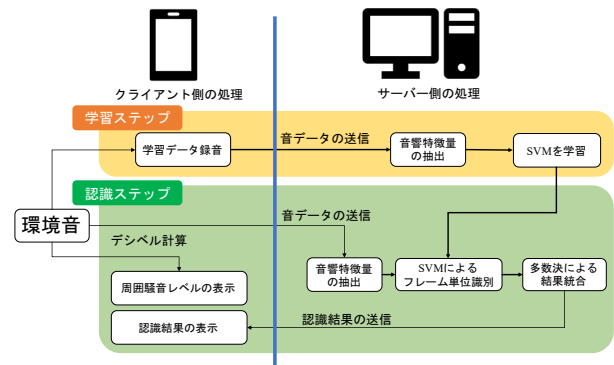


図 1 システム構成図

Fig. 1 System configuration

どの特別な機器や高度な認識手法を用いずとも高い認識性能を実現することができた。

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、環境音の収録および情報の表示をクライアント側のタブレット端末 (iPad 9.7 インチ) で、環境音の認識処理をサーバー側で行う。

このようなクライアントサーバーシステムを今回採用したのは、以下のような理由からである。

- 端末自体の高度な計算能力を必要としないため、一般的なスマートフォンやタブレットに加え、AR グラスやスマートウォッチなど小型のウェアラブルデバイスでも利用可能な構成だと考えられる。
- 将来的に、深層学習などの最先端の認識手法を、常にシステムに適用することができる。

次に、環境音認識を行う際の処理を説明する。まず学習ステップとして、クライアント側の端末内蔵マイクで録音した数サンプル (あるいは数十秒) の音データをサーバー側に送信し、音響特徴量の抽出を行なった後に識別器 (今回は SVM) の学習を行う。一方、認識ステップでは、クライアント側の端末内蔵マイクで録音した音データをサーバー側に連続的に送信し続ける。サーバー側では、音響特徴量の抽出や結果統合といった処理を行なった後、認識結果をクライアント側に送信する。認識結果を受け取るクライアント側の端末は、結果に対応した情報を実時間で表示する。

また、環境音認識に併せて、周囲の騒音レベルの表示を行なっている。こちらの処理はクライアント側の端末内で行なっており、サーバー側との通信は必要としていない。

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 として用いた^[27]。先行研究^[16-24]では、音響特徴量として Mel-frequency cepstral coefficients

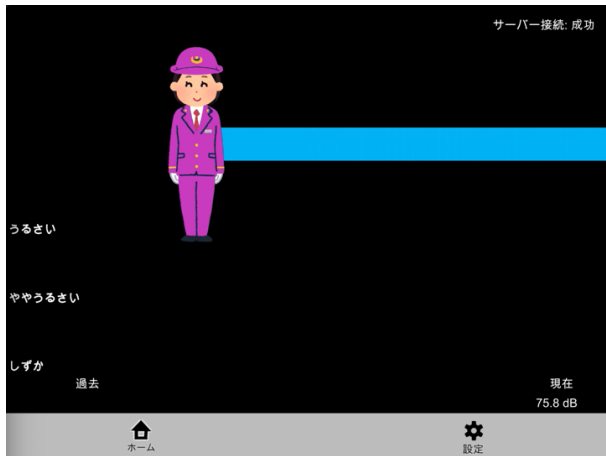


図 2 システムの画面例：環境音認識のみを行う場合（女声アナウンスの場合）

Fig. 2 System screen example: only environmental sound recognition (In case of female voice announcement)

(MFCC) や PLP が多く用いられている。また、音声認識においては MFCC よりも PLP の方が音響的なミスマッチに頑健であるとの報告がある [28]。我々が先に行なった室内空間における環境音認識実験では、学習ステップと認識ステップで音データの収録状況に差異が存在していた。そのような学習データとテストデータの音響ミスマッチが大きくなる状況において、MFCC よりも PLP が有効であることを確認している [25]。

3.2 識別器について

提案するシステムでは、端末の内蔵マイクで認識したい音を事前に録音するという作業を行うことから、必然的に学習データは少量になると予想される。そこで識別器には、少ない学習データで動作し、かつ利用環境の変動に頑健である Support Vector Machine (SVM) を用い、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。

3.3 フレーム単位識別結果のスムージングについて

識別器によりフレーム単位で認識された結果に対し、0.5 秒の範囲内で多数決を行い、その区間の認識結果とするスムージングを行う。なお、スムージング区間は 0.5 秒の間隔でシフトするため、スムージング区間毎に重畳している部分は存在しない。また、多数決の結果において同率のクラスが複数ある場合は、1 つ前のスムージング区間の結果を利用する。

3.4 環境音可視化システムの実装および画面設計について

環境音を可視化するクライアント側のアプリケーションは Unity [29] を用いて開発した。今回はタブレットに iPad を利用したため、基本的には iOS 向けのアプリケーション

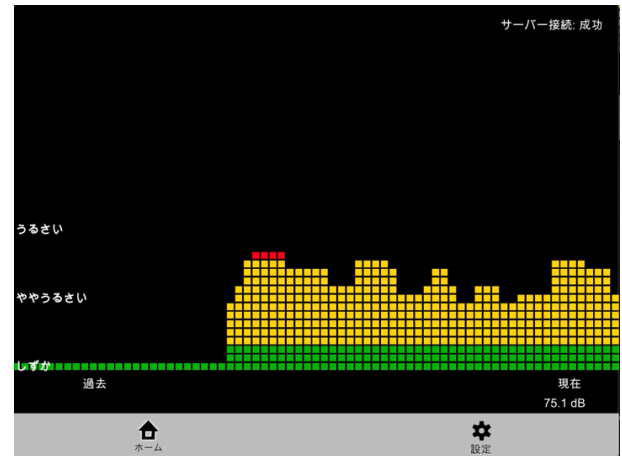


図 3 システムの画面例：騒音レベル表示のみを行う場合

Fig. 3 System screen example: only noise level

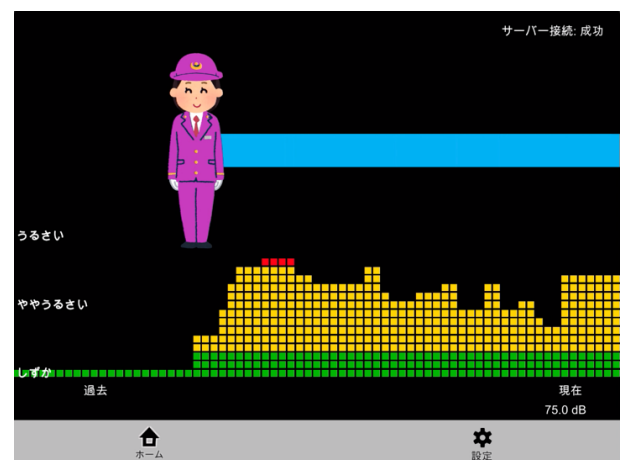


図 4 システムの画面例：環境音認識および騒音レベル表示の両方を行う場合（女声アナウンスの場合）

Fig. 4 System screen example: environmental sound recognition and noise level (In case of female voice announcement)

とした。しかし、Unity はマルチプラットフォームに対応した言語のため、Windows や macOS, Android といった異なる OS 用アプリケーションに即座に切り替えることができ、様々なデバイス上で利用することが可能である。

図 2, 図 3, 図 4 に今回開発した環境音可視化システムの 3 種類の情報提示方法を示す。4.3 節で詳述するが、図 2 は環境音認識結果のみを提示する方法、図 3 は騒音レベルのみを提示する方法、図 4 は環境音認識結果と騒音レベルの両方を提示する方法である。

4. 実験

本章では実験の概要および結果について述べる。最初に、システムが認識すべき音の選定のため、5 名の健聴者を対象に予備実験を行なった。その後、この予備実験の結果を元にシステムを作成し、図 2, 図 3, 図 4 に示した 3 種類の情報提示方法の中で最も有効なものを調査するため、主観

表 1 収録環境毎に選定した認識対象音

Table 1 Selected sounds for each recording environment

収録環境名称	認識対象音
バス車内	車内アナウンス (女声の録音) IC カードの接触音 ドア開閉時のブザー
電車車内	車内アナウンス (男声, 女声) ドア開閉時のブザー 発車前のメロディ
駅構内	女声アナウンス (上り線) 男声アナウンス (下り線) 発車前のメロディ

評価実験を 12 名の健聴者を対象に実施した。最後に、提案手法を用いたシステムの認識性能評価実験を行なった。

4.1 実験データの概要

本報告ではデータの収録環境を公共交通機関に絞り、路線バスの車内、電車の車内（東急田園都市線、横浜市営地下鉄ブルーライン）、駅構内（東京メトロ永田町駅）の 3 種類の環境でデータの収録を行なった。3 章および図 1 では環境音の収録はクライアント側のタブレット端末で行うとしているが、予備実験および主観評価実験では動画データが必要となるため、人間の視野角に近い広角動画の撮影が可能な GoPro^[30] を用いて、音声付き動画を撮影した。識別器の学習データには、撮影された動画の音データを、波形の視察によって切り出して利用した。録音データの形式は Linear PCM 形式、モノラル録音、サンプリング周波数 16,000 Hz、量子化 16 bit とした。また、音データから PLP を抽出する際は、フレーム長 60 ms、フレームシフト幅 10 ms で行なった。

4.2 予備実験

環境音には多様な種類が存在している。提案システムの主観評価や認識性能評価を行うにあたり、健聴者 5 名を対象に予備実験を実施し、各状況下でシステムが認識すべき音を選定した。

予備実験の手順を説明する。まず、バス車内、電車車内、駅構内の動画を視聴してもらい、動画中に存在していた環境音を列挙してもらった。その後、列挙した環境音の中で、自分が重要だと思う音を 3 つ選択し、理由と併せて答えてもらった。この予備実験によって選定されたシステムの認識対象音の一覧を表 1 に示す。

4.3 主観評価実験

今回、システムが音環境情報を提示する方法として、

- (1) 対象音を認識すると、音の検出時にアイコン画像で音の種類を示し、その後、青色のバーで音の継続を表示する方法 (図 2)

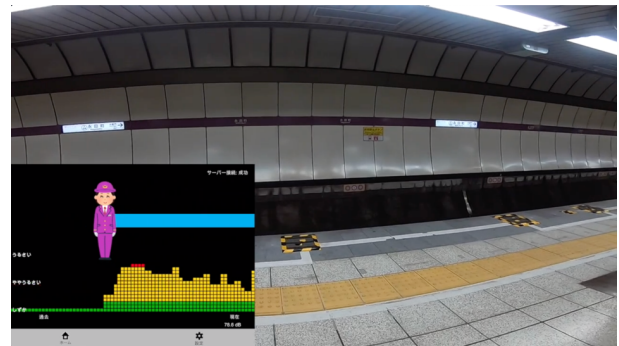


図 5 主観評価実験で用いた動画の例

Fig. 5 Example of video used in subjective evaluation experiment

- (2) 音の存在を視覚的に再現した、騒音レベルを表示する方法 (図 3)

- (3) 上記 2 つを同時に表示する方法 (図 4)

の 3 種類の情報提示方法を検討した。特に (2) については、振動等を用いて音の存在を提示する聴覚障害支援デバイス^[31] と同等の情報を視覚的に表示するものとして検討した。この 3 種類の情報提示方法の音環境の理解し易さ（以下、音環境理解度）を調査するため、健聴者 12 名を対象に主観評価実験を実施した。なお主観評価実験はあくまで有効な情報提示方法を調査するのが目的のため、システムの環境音認識部分には視察で求めた正解データを表示した。

主観評価実験の手順を説明する。被験者には、バス車内、電車車内、駅構内の 3 種類の環境で撮影された動画（以下、シチュエーション動画）を視聴してもらう。各シチュエーション動画の長さはそれぞれ 5 分程度であり、4.2 章で選定した環境音が必ず存在するよう編集した。また、シチュエーション動画の左下には、3 種類の情報提示方法の内の 1 種類を実時間で表示した。すなわち、3 種類の環境のシチュエーション動画に対して、3 種類の情報提示方法のシステムがあるため、被験者は合計 9 本の動画を視聴している。主観評価実験で被験者が視聴した動画の例を図 5 に示す。

9 本の各動画の視聴終了毎に、以下の 2 つの設問に答えてもらった。

- 動画左下に表示されていたシステムを利用することで、どの程度シチュエーション動画内の音環境を理解できるかを、5 段階評価でお答えください。
- 動画を視聴し、システムの提示方法についての感想をご自由にお書きください。

また、9 本全ての動画を視聴した後に以下の設問に答えてもらい、全体評価とした。

- 「環境音認識のみ」「騒音レベルのみ」「環境音認識+騒音レベル」の 3 種類のシステムの情報提示方法の中で、どれが一番音環境を理解できると感じましたか。情報提示方法毎の音環境理解度の結果を図 6 に示す。図

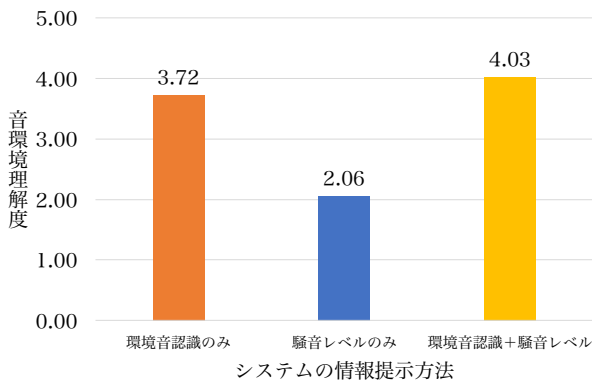


図 6 情報提示方法毎の音環境理解度の結果

Fig. 6 Results of environmental sound understanding level for each visualization method

表 2 全体評価の結果

Table 2 Results of overall evaluation

システムの情報提示方法	回答人数
環境音認識のみ	4
騒音レベルのみ	0
環境音認識+騒音レベル	8

中の棒グラフの数値はそれぞれ、情報提示方法毎の音環境理解度を平均した値である。

また、全体評価の結果を表 2 に示す。これらの結果から、今回のように視覚的に情報を表示する場合、騒音レベルのみを提示する方法ではあまり有益な情報が得られていないことが分かった。一方で、環境音認識が完全に動作すると仮定した場合、音の種類とその継続長を伝えることの有用性の高さが示唆された。そして、騒音レベルと環境音認識を併用した場合に一番有用性が高くなるというのが、図 6 と表 2 から見て取れる。自由記述式の設問の回答においても、騒音レベルと環境音認識を併用することで、対象音あるいはそれ以外の音の有無、加えてその環境の騒々しさが分かり易かったという旨のコメントが多く見られた。

4.4 認識性能評価実験

本節では、提案手法を用いた環境音可視化システムの認識性能を評価し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。認識性能の評価に用いた音データは、4.1 節で述べたように GoPro で撮影された動画より抽出した。各収録環境の学習データとテストデータに用いた音データの時間長を表 3 に示す。識別器の学習は収録環境毎に行なっているため、今回 4 つの学習モデルを作成した。また、各収録環境において学習データとテストデータは同じ環境で収録されているが、別日に収録あるいは収録位置を変えるなど、収録状況としてはオープン評価となるようにした。なお、表 3 中の「電車車内 1」は東急田園都市線の車内、「電車車内 2」は横浜市営地下鉄ブルーラインの車内の環境を示す。

表 3 実験に用いた音データの時間長 (単位: 秒)

Table 3 Amount of sound data (Unit: sec)

収録環境	認識対象音	学習データ	テストデータ
バス車内	車内アナウンス	34.0	83.6
	IC カードの接触音	5.0	6.8
	ドア開閉時のブザー	12.1	18.7
	その他 (雑音区間)	62.1	191.1
電車車内 1	車内アナウンス	16.0	47.0
	ドア開閉時のブザー	5.0	15.0
	発車前のメロディ	3.9	4.7
	その他 (雑音区間)	52.1	150.9
電車車内 2	車内アナウンス	20.7	128.6
	ドア開閉時のブザー	4.8	14.0
	発車前のメロディ	4.0	11.6
	その他 (雑音区間)	82.5	228.9
駅構内	女声アナウンス	14.8	59.7
	男声アナウンス	14.3	90.0
	発車前のメロディ	12.3	12.6
	その他 (雑音区間)	42.1	138.2

各収録環境における、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、収録環境毎に各認識対象音の F-measure を平均した値である。また、各収録環境における認識対象音ごとのフレーム単位識別の結果を図 8 に示す。

まず、車内の環境 (バス車内、電車車内 1、電車車内 2) に関して、車内アナウンスの認識率が平均的に低い傾向にあると分かる。これは車両走行音などの雑音によって、認識対象音の SNR が低下していることが要因であると考えられる。その一方で、ドア開閉時のブザーは必ず停車中に行われるため雑音が少なく、車内アナウンスに比べると平均的に認識率は高い傾向にあった。停車中に発生する音として発車前メロディも挙げられるが、電車車内 2 の発車前メロディは F-measure 0.52 と上手く認識できていないことが見て取れる。一方で、電車車内 1 (東急田園都市線) の発車メロディは F-measure 0.88 と高い性能であった。これは、電車車内 2 の発車前メロディが音楽的なもの、かつ SNR が 10.5 dB 程度であったのに対し、電車車内 1 の発車前メロディは単調で連続的なブープ音、かつ SNR が 15.8 dB であったためと考えられる。加えて、駅構内における発車前メロディの認識率は電車車内 1 および電車車内 2 のものと比べると著しく高い。これは、音源が車内から間接音として録音されたものではなく、マイクと音源の間に遮蔽物がない状態で発車前メロディを録音できていたために高い認識性能に繋がったと考えられる。

一方、実際にこのシステムを使用する場合、利用する状況あるいは対象とする音によって、要求される認識性能が異なってくることが予想される。例えば、車のクラクションや、避難を促すサイレン・アナウンスなど、危険性あるいは緊急性が高い音を対象とした場合、現状のシステムの認識性能ではまだまだ改善が必要であると言える。一方で、

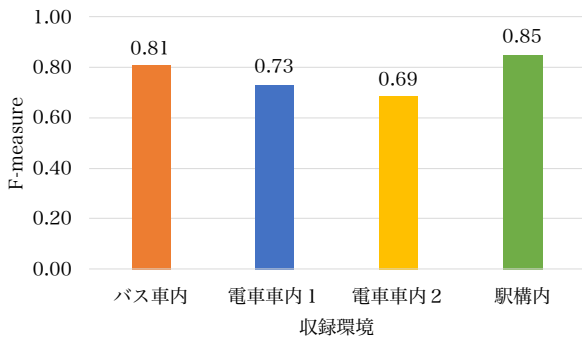


図 7 フレーム単位による認識精度の結果（収録環境ごとの平均）

Fig. 7 Results of recognition accuracy by sound frame (Average of each recording environment)

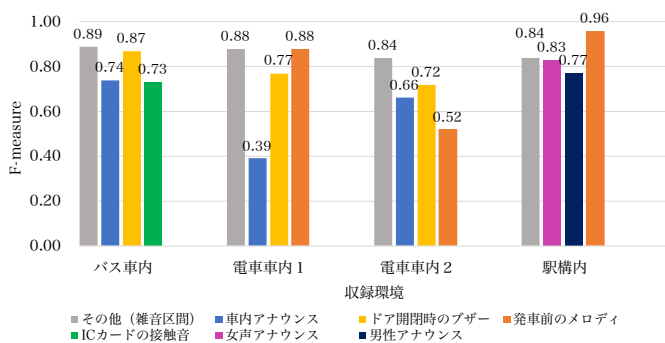


図 8 フレーム単位による認識精度の結果（認識対象音ごとに表示）

Fig. 8 Results of recognition accuracy by sound frame (Each target sound)

今回対象としたようなドア開閉時のブザーや、次のバス停名を告げる車内アナウンスなど、危険性が少ないあるいは後ほど視覚的に見ても分かるような状況下では、現状の認識性能でも聴覚障害者の支援となり得る可能性があると考えている。

次に、今回の認識性能のシステムが、4.3 節で用いた完全な精度で動作するシステムと比べた場合に、音環境理解度にどの程度の影響を与えるのかについて調査した。表 2 で最も回答人数が多かった「環境音認識+騒音レベル」による情報提示方法を採用し、健聴者 5 名に対して実際のシステムの認識性能による主観評価実験を行なった。この主観評価実験の手順を説明する。被験者には、4.3 節と同様に、各シチュエーション動画を視聴してもらう。各シチュエーション動画の左下には、「環境音認識+騒音レベル」の情報提示方法により、実際の認識性能のシステムを実時間で表示した。結果を図 9 に示す。両者の音響理解度にあまり大きな差はないことが分かる。これはアナウンスや発車前メロディなど、その音の継続を全て認識できていなくても、途切れ途切れの認識ができてさえいれば、ある程度状況が推測できるということを示唆している。

最後に、音の事前登録を行う提案手法の有効性を確認するため、インターネット上で利用可能な巨大な音イベント

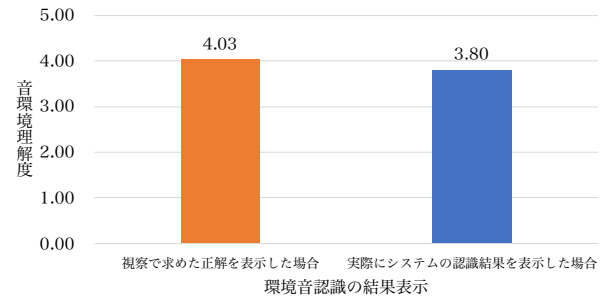


図 9 本報告における環境音認識性能のシステムを用いた場合の音響理解度の調査

Fig. 9 Results of comparative experiments on environmental sound understanding level between oracle recognition performance and actual recognition performance

表 4 AudioSet より用いた音データの時間長（単位：秒）

Table 4 Amount of sound data derived from the AudioSet (Unit: sec)

認識対象音	オントロジー名称	学習データ
車内アナウンス (女声の録音)	"Female speech, woman speaking"	524.1
IC カードの接触音	"Beep, bleep"	477.6
ドア開閉時のブザー	"Buzzer"	521.0
その他 (雑音区間)	"motor vehicle", "public space", "silence"	1679.8

データセットを用いて識別器の事前学習を行なった場合との性能を比較した。このデータセットとしては、Google が公開している環境音データベース (AudioSet^[32]) を用いた。表 4 に、各認識対象音の学習に利用した音データの AudioSet オントロジー名称と時間長を示す。AudioSet オントロジーとは、音の階層構造の集合であり、人間や動物の音、自然音や環境音、音楽や雑音など、幅広い日常音をカバーしている。AudioSet オントロジーの構築については Gemmeke らの論文^[32] で詳しく述べられている。ここでは、表 3 にあるバス車内の学習データで学習したモデル (以下、カスタムモデル) と、表 4 に示した AudioSet データで学習したモデル (以下、AudioSet モデル) の認識性能を比較するため、表 3 のバス車内のテストデータを対象に認識実験を行なった。なお、どちらの学習モデルも識別器に 3.2 節で述べた SVM を用いている。結果を図 10 に示す。図中の棒グラフの数値はそれぞれ、バス車内のテストデータを対象にそれぞれの学習モデルで認識を行い、各認識対象音の F-measure を平均した値である。AudioSet モデルは学習に利用したデータサイズがおおよそ 1 時間程度と大きいにも関わらず、F-measure 0.25 とカスタムモデルよりはるかに低い性能となった。一方、カスタムモデルは学習におおよそ 2 分程度のデータを使ったにすぎないが F-measure 0.81 を示した。このように、聴覚障害者が日頃から立ち寄る環境で認識対象音を事前に登録するというカスタマイズを行うことで、一般的な環境音認識システムでは達成できないレベルの認識精度を簡便なシステムでも実現できることが示唆された。

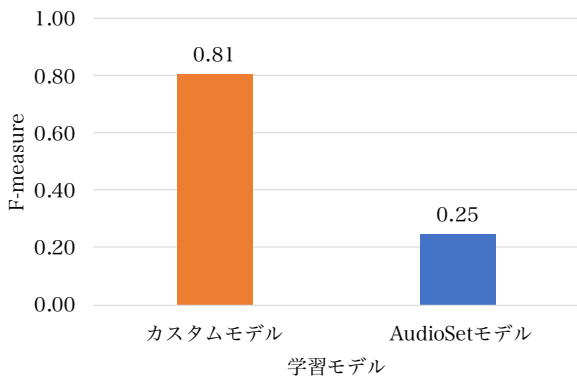


図 10 カスタムモデルと AudioSet モデルの認識性能の比較結果

Fig. 10 Results of comparative experiments on recognition performance between proposed model and AudioSet model

5. おわりに

本報告では、聴覚障害者が日頃立ち寄る場所において、事前に対象音をシステムに登録するという方式の環境音可視化システムを提案した。また、その有用性を検討するため、タブレット上にシステムを実装し実験を行なった。健聴者を対象として行った主観評価実験では、騒音レベルと環境音認識を併用した情報提示方法の有効性を確認した。また、認識性能評価実験では、数秒から数十秒といった少量の学習データで平均 F-measure 0.77 を得ることができた。加えて、巨大なデータベースを用いて一般的な環境音モデルを構築した場合との性能比較を行い、認識対象音を事前に登録しておく方法の有効性を確認した。

今後、提案システムを実用化するには、まだ多くの課題が残されている。まず、現状は複数音源を同時に認識することはできないこと。そして、現状では識別器の学習に用いるデータを手動で切り出しているため、システムを実用化するには、録音データに対して音区間検出を行い、学習データに用いる音クリップを簡便に抽出できる仕組みが必要であることなどが挙げられる。さらに、聴覚障害者がシステムを日常的に利用するには、スマートウォッチや AR グラスといったデバイスに情報を通知する必要も考えられる。今後、これらの課題に対処していくことで、聴覚障害者の方々から本当に「使ってみたい」と感じてもらえるような、実用性の高い聴覚障害支援システムを実現したい。

謝辞 本研究を行うにあたり、筑波技術大学産業技術学部 平賀瑠美先生、筑波技術大学 加藤優さん、設楽明寿さん、中原夕夏さん、小川直希さんから有益な助言をいただきました。

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] 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.

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

[4] 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.

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

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 根岸 佑也, 河口 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.
- [20] 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.
 - [21] 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.
 - [22] I. Trancoso, J. Portelo, M. Bugalho, J. Neto, A. Serralheiro, “Training audio events detectors with a sound effects corpus”, Proc. Interspeech, 2008.
 - [23] J. Rouas, J. Louradour, S. Ambellouis, “Audio Events Detection in Public Transport Vehicle”, Intelligent Transportation Systems Conference, 2006.
 - [24] M. Bugalho, J. Portelo, I. Trancoso, T. Pellegrini, A. Abad, “Detecting Audio Events for Semantic Video Search”, Proc. Interspeech, pp. 1151-1154, 2009.
 - [25] 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”, 研究報告アクセシビリティ (AAC-5), No. 11, pp. 1-6, 2017.
 - [26] 浅井研哉, 志磨村早紀, 北義子, 綱川隆司, 西田昌史, 西村雅史, “聴覚障害者支援のための実環境下における環境音認識”, 情報処理学会第 80 回全国大会, 1ZB-04, pp.519-520, 2018.
 - [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.
 - [29] Unity (<https://unity3d.com/jp>), (参照 2019-2-1).
 - [30] GoPro (<https://jp.shop.gopro.com/cameras>), (参照 2019-2-1).
 - [31] Antenna - 髪の毛で音を感じる新しいユーザーインタフェース (<http://antenna.jp/>), (参照 2019-2-1).
 - [32] J. F. Gemmeke, D. P. W. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, M. Ritter, “Audio Set: An ontology and human-labeled dataset for audio events”, ICASSP 2017, pp.776-780, 2017.