

# 聴覚障害者支援のための実環境下における環境音認識

浅井 研哉<sup>†</sup> 志磨村 早紀<sup>‡</sup> 北 義子<sup>§</sup> 綱川 隆司<sup>†</sup> 西田 昌史<sup>†</sup> 西村 雅史<sup>†</sup>

<sup>†</sup>静岡大学大学院 <sup>‡</sup>早稲田大学障がい学生支援室 <sup>§</sup>国立障害者リハビリテーションセンター学院

## 1. はじめに

聴覚障害者は会話などの音声に加えて、生活環境音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が、聴導犬は普及が困難となっている.

そこで我々は、居住空間内における生活環境音をシステムにより認識し、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についての研究を行なっている<sup>[2]</sup>. 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の向上に繋げることも期待出来る. 本研究では、実環境下における環境音認識システムの性能改善に関する検討を行なった.

## 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>[3]</sup>と、深層学習を用いないもの<sup>[4][5]</sup>に分けることができ、研究の幅は広い. しかし、本報告が提案す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あまりない.

## 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、対象音を登録するというシステムを想定している.

音響特徴量には Perceptual Linear Prediction (PLP) を次数 24, フレーム長 60 ms, フレームシフト幅 10 ms として用いる. 実際のシステム利用時には、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.

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、様々な環境に汎用的に対応するため特

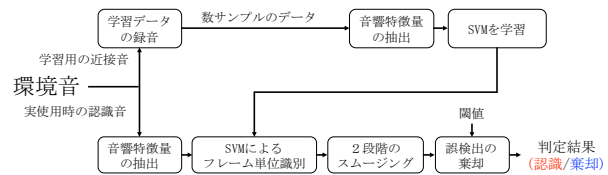


図1 システム構成図

にパラメータチューニングは行っていない. フレーム単位の識別結果は2段階のスムージング処理 (1段階目:  $i$  フレーム目の前後 50 フレームで多数決  $\rightarrow i$  フレーム目の結果とする, 2段階目:  $i$  フレーム目の前後 25 フレームで多数決  $\rightarrow i-25 \sim i+25$  フレーム目の結果とする) を適用する<sup>[2]</sup>.

最後に、誤検出の棄却を行う. スムージングされた結果を用いて、ある対象音として推定された区間におけるフレーム毎のクラス所属確率を平均する. その後、その平均値が閾値より高ければ認識し、低ければ棄却を行う. なお、閾値は環境に合わせてユーザーが自由に調節することを想定している.

## 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、聴導犬が通知を行なっている音を基準に、認識対象とする環境音の選定を行なった. 今回は、ヤカン沸騰時の警笛音 (boiling), ドアベル (doorbell), 電子レンジの終了音 (microwave), 着信音 (ringtone), 目覚まし時計のアラーム (alarm) の5つを認識対象音とした.

### 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、3種類の実環境下にて音データの録音を行なった. そのうち、評価用に用いた実験環境 (a), (b) のレイアウト図をそれぞれ図2, 図3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.

マンションである実験環境 (a) 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.

オフィスである実験環境 (b) についても、換気扇・空調を動作させたまま録音を行い、在室し

Environmental Sound Recognition in a Real Environment for Hearing Impaired

<sup>†</sup>Kenya Asai, <sup>‡</sup>Saki Shimamura, <sup>§</sup>Yoshiko Kita, <sup>†</sup>Takashi Tsunakawa, <sup>†</sup>Masafumi Nishida, <sup>†</sup>Masafumi Nishimura

<sup>†</sup>Shizuoka University

<sup>‡</sup>Waseda University Office for Students with Disabilities

<sup>§</sup>National Rehabilitation Center for Persons with Disabilities

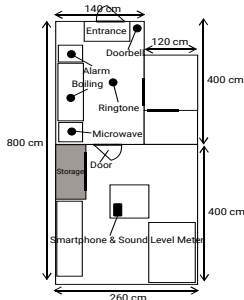


図 2 実験環境(a)：各音源と測定位置，ドアは締め切った状態で収録

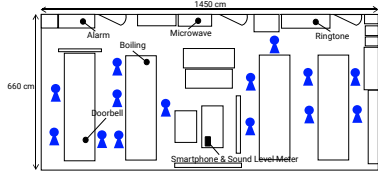


図 3 実験環境(b)：各音源と測定位置，及び実験中に在室していた人が主に居た位置を示す

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(b)においては話し声や物音（キーボードのタイプ音，ドアの開閉音等）といった雑音が多い環境下になっている。

また、どの実験環境のテストデータも外部からの雑音（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 実験データの詳細

音データは Linear PCM 形式，モノラル，サンプリング周波数 48,000 Hz，量子化 16 bit で収録した。また、雑音クラスの学習には、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 1 に示す。実験データの内、実験環境(a)，(b)のデータは評価用データセットとして用いた。これら 2 種類の環境とは別のマンションで同様な録音を行なった実験環境(dev.)のデータは、誤検出棄却を行う際の閾値設定用に用いた。

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用いた。テストデータには、実験環境の定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。

## 5. 実験結果

評価指標には、適合率と再現率の調和平均で

表 1 実験データの時間長 (単位: 秒)

データ	学習データ			テストデータ			
	(a)	(b)	(dev.)	(a)	(b)	(dev.)	
実験環境							
対象音	boiling	12.3	11.3	10.9	175.3	212.8	189.3
	doorbell	9.8	9.8	9.8	30.0	40.0	40.0
	microwave	3.2	4.8	14.8	54.0	16.0	60.0
	ringtone	2.7	3.3	3.2	14.0	14.0	14.0
	alarm	12.3	12.3	12.2	50.0	50.0	49.0
	other	10.0	10.0	10.0	3119.2	3183.9	3120.4

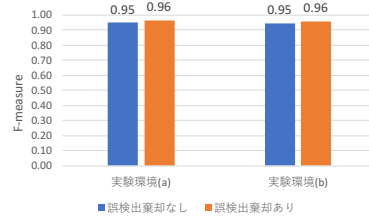


図 4 イベント単位による認識精度の結果

表 2 イベント単位識別時の実験結果の内訳

実験環境	対象音	検出回数	誤検出棄却なし		誤検出棄却あり	
			誤検出回数	誤検出の原因	検出回数	誤検出回数
(a)	boiling	20 / 20	0		20 / 20	0
	doorbell	19 / 20	0		18 / 20	0
	microwave	20 / 20	1	alarmの余韻:1回	19 / 20	0
	ringtone	20 / 20	0		16 / 20	0
	alarm	20 / 20	11	boiling:10回, doorbell:1回	20 / 20	0
	boiling	20 / 20	7	ドアの開閉音:7回	20 / 20	5
(b)	doorbell	20 / 20	0		20 / 20	0
	microwave	18 / 20	2	ドアの開閉音:2回	18 / 20	0
	ringtone	20 / 20	0		18 / 20	0
	alarm	20 / 20	1	ドアの開閉音:1回	20 / 20	0

ある F-measure を用いた。図 4 に、各実験環境におけるイベント単位による認識結果を示す。

図中の棒グラフの数値はそれぞれ、実験環境毎(ringtone, alarm)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、誤検出棄却を適用した場合に精度が改善した。

また、表 2 に認識結果の詳細な内訳を示す。全ての実験環境で、誤検出の棄却を行った場合に結果が改善しており、特に実験環境(a)においては顕著な改善が見られた。一方で、実験環境(a)の ringtone のように、再現率の低下もあった。

## 6. おわりに

今後は、意図的に雑音を発生させた環境下で環境音認識実験を行い、そのような状況下での性能改善について検討する。

## 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] 浅井研哉 他, “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, Vol.2017-AAC-5, No.11, pp.1-6, 2017.
- [3] O.Gencoglu et al.: “Recognition of acoustic events using deep neural networks”, European Signal Processing Conference, 2014.
- [4] A.Pillos et al.: “A REAL-TIME ENVIRONMENTAL SOUND RECOGNITION SYSTEM FOR THE ANDROID OS”, Detection and Classification of Acoustic Scenes and Events, 2016.
- [5] M.Bugalho et al.: “Detecting Audio Events for Semantic Video Search”, Proc. Interspeech, pp. 1151-1154, 2009.