5ZB-07

実環境を想定したスマートグラスを用いた日常生活音の認識

平井　和之　　西田　昌史　　綱川　隆司

静岡大学

1.　はじめに

聴覚障がい者は耳が聞こえないため、周囲の状況の把握が困難である。聴覚障がい者の環境音認知の向上のために視覚情報を利用した環境音学習の有効性について検討されている[1]。環境音の中から緊急の回避や避難を必要とする警告音を識別し、ウェアラブル端末やスマートフォンへの画面表示をすることで、聴覚障がい者の日常生活を支援するという研究も行われている[2]。また、環境音を認識し、スマートフォンやタブレットなどの端末に表示させることで環境音を可視化するという研究も行われている[3]。さらに、スマートグラスを用いて、人の音声や環境音を認識し、可視化するシステムの研究が行われている[4]。文献[4]の可視化システムの認識性能を向上させる研究も行われている[5]。

これまでのスマートグラスを用いた可視化システムは、オフラインでの処理を行っており、また認識対象の環境音のみが鳴っている状況を想定したシステムである。そこで、本研究ではオンライン処理と雑音を考慮した実環境を想定した日常生活音の認識について検討した。

2.　可視化システム

本研究では、環境音の録音および可視化に用いるスマートグラスとして、図 1 の EPSON MOVERIO BP-300 を使用する。

図 1　EPSON MOVERIO BP-300

可視化システムの概要を図 2 に示す。スマートグラスをクライアント側として、クライアント側で環境音を録音し、録音した環境音をサーバー側で認識し、その認識結果をクライアント側に送って、スマーグラスで可視化する。

従来の可視化システムでは認識手法に CatBoost を用いている。環境音はサンプリング周波数 16kHz で収集し、CatBoost を用いる際の特徴量抽出ではフレーム長 25ms、シフト幅 10ms で抽出した MFCC24 次元を用いていた。

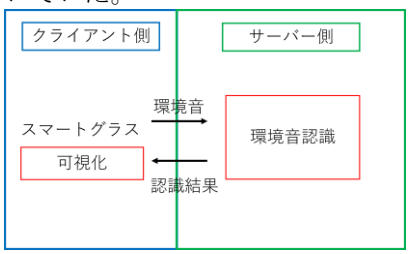

図 2 可視化システム概要

3.　提案手法

本研究では、認識対象の環境音が連続して鳴っている状況における実時間認識手法としてフレームの概念を利用した。図 3 のようにフレーム長 0.5s のフレームをシフト幅 0.25s ずつずらしながら認識と可視化を繰り返すことで、実時間認識に加え、認識漏れの防止が可能となる。

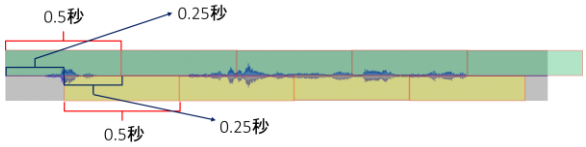

図 3　実時間認識手法イメージ図

また、日常生活の雑音環境下に耐える頑健性の高い認識手法として LSTM・GRU を利用する。また、従来研究では MFCC24 次元を用いているが、環境音認識において他に用いられる特徴量として LFCC、openSMILE、MFB(メルフィルタバンク)が存在する。本研究では、SVM、CatBoost、LSTM、GRU の 4 つの認識手法と MFCC24 次元、LFCC24 次元、openSMILE26 次元、MFB128 次元の 4 つの特徴量の組合せで、より頑健性の高い組合せを調査する。

4.　評価実験

認識手法には SVM、CatBoost、LSTM、GRU を使用し、各認識手法における特徴量として MFCC、LFCC、openSMILE、MFB を使用して環境音認識性能の評価実験を行った。

評価実験に用いる環境音 18 種類を表 1 に示す。表 1 においてテレビの音(会話のみ、音楽のみ、会話と音楽のみ)を「その他」としている。

表 1　環境音の種類

お風呂の通知音	インターホンの音	ケトルの音	救急車のサイレン音	ドアの閉まる音	レンジの音	体温計の音
冷蔵庫の通知音	緊急地震速報の通知音	携帯の着信音	洗濯機の通知音	炊飯器の通知音	目覚ましのアラーム音	その他
掃除機の音	赤ちゃんの泣き声	家族の呼びかけ	無音			

表 1 の環境音 18 種類のうち「その他」以外を 60 個ずつ、「その他」を 180 個、計 1200 個使用した。環境音データの長さはすべて 0.5 秒である。また、

Daily life sound recognition using smart glasses for real environments
Kazuyuki Hirai, Masafumi Nishida, Takashi Tsunakawa
Shizuoka University

Copyright ©2024 Information Processing Society of Japan.
All Rights Reserved.

雑音環境下を疑似的に再現するために、同じく 0.5 秒の雑音データを 1200 個すべてに合成した。使用した雑音データは屋内で聞こえる屋外のエンジン音などを雑音としている。実験条件は以下の 3 つに分けた。

① 雑音を含まない場合
② 学習データ・テストデータ両方に雑音を含む場合
③ テストデータのみに雑音を含む場合

それぞれの実験条件において、学習データは「その他」以外を各 50 個、「その他」を 150 個計 1000 個使用し、テストデータは「その他」以外を各 10 個、「その他」を 30 個計 200 個使用した。また、その場合にできる 6 パターンで学習とテストを行い、テストデータがどの程度正しく認識されるか交差検証を行った。以下の表 2、表 3、表 4 に各実験条件の実験結果を示す。認識性能には F 値を用いており、6 パターンそれぞれの認識性能の平均であり、F 値はイベント単位である。また、RTF(Real Time Factor)はテストデータ 1 個当たりの RTF の平均である。また、RTF は 1 より小さければ実時間処理が可能であることを示す指標である。

表 2　雑音を含まない場合の実験結果

		SVM	CatBoost	LSTM	GRU
認識精度(F 値)	MFCC	0.97	0.99	0.97	0.98
	MFB	0.68	0.99	0.97	0.97
	LFCC	0.97	0.98	0.98	0.98
	Opensmile	0.97	0.99	0.99	0.99
RTF	MFCC	0.195	0.007	0.345	0.345
	MFB	0.641	0.049	0.235	0.232
	LFCC	0.135	0.005	0.261	0.276
	Opensmile	0.317	0.159	0.346	0.338

表 3　学習データ・テストデータ両方に雑音を含む場合の実験結果

		SVM	CatBoost	LSTM	GRU
認識精度(F 値)	MFCC	0.92	0.96	0.97	0.97
	MFB	0.68	0.99	0.94	0.94
	LFCC	0.91	0.97	0.95	0.96
	Opensmile	0.95	0.97	0.97	0.97
RTF	MFCC	0.261	0.007	0.340	0.317
	MFB	0.631	0.060	0.212	0.217
	LFCC	0.177	0.004	0.253	0.240
	Opensmile	0.353	0.161	0.329	0.336

表 4　テストデータのみに雑音を含む場合の実験結果

		SVM	CatBoost	LSTM	GRU
認識精度(F 値)	MFCC	0.71	0.71	0.76	0.77
	MFB	0.68	0.73	0.85	0.85
	LFCC	0.78	0.82	0.79	0.79
	Opensmile	0.85	0.86	0.83	0.82
RTF	MFCC	0.193	0.031	0.314	0.323
	MFB	0.634	0.054	0.231	0.256
	LFCC	0.136	0.005	0.263	0.240
	Opensmile	0.316	0.162	0.338	0.371

表 2、表 3 より雑音を含む場合と学習・テストデータ両方に雑音を含む場合では認識手法と特徴量の組合せによって認識性能に大きな違いは見られなかった。表 4 より、CatBoost と openSMILE の組合せが最も認識精度が高かった。この組合せは表 2、表 3 においても高い認識精度であった。また、どの手法においても RTF は 1 より小さいので実時間処理が可能であることが分かった。以下の表 5 に③の実験条件における CatBoost と openSMILE の組合せの混同行列を示す。なお、表 5 において列は予測結果を示し、行は正解を示している。

表 5　③の CatBoost・openSMILE での混同行列

	お風呂	インターホン	ケトル	サイレン	ドア	レンジ	体温計	冷蔵庫	地震速報	携帯	洗濯機	炊飯器	目覚まし	赤ちゃん	掃除機	呼びかけ	無音	その他
お風呂	51	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	7
インターホン	0	58	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
ケトル	0	0	41	0	0	0	0	0	0	0	0	0	0	0	0	17	0	2
サイレン	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ドア	0	0	0	0	59	0	0	0	0	0	0	0	0	0	0	0	0	1
レンジ	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0
体温計	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	10
冷蔵庫	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0
地震速報	0	0	0	0	0	0	0	0	31	0	0	0	0	0	0	29	0	0
携帯	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0
洗濯機	0	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0
炊飯器	0	0	0	0	0	0	0	0	0	0	0	59	0	0	0	1	0	0
目覚まし	0	0	0	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0
赤ちゃん	0	0	0	1	0	0	0	0	1	0	0	0	0	35	1	6	0	16
掃除機	0	0	1	0	0	0	0	0	1	0	0	0	0	0	57	1	0	0
呼びかけ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60	0	0
無音	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	56	1	1
その他	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	172

表 5 より、「ケトル」と「地震速報」と「無音」は「呼びかけ」に、「体温計」と「赤ちゃん」は「その他」に誤認識することが多かった。

5.　おわりに

本研究では、オンライン処理における実時間認識手法に加え、雑音環境下での日常生活音の認識でより高い認識性能を確保するために SVM・CatBoost・LSTM・GRU の 4 つの認識手法と MFCC・LFCC・openSMILE・MFB の 4 つの特徴量の組合せを提案した。評価実験の結果、CatBoost と openSMILE が雑音環境下で頑健性の高い組合せであることが分かった。この評価実験はプログラム上で行ったものであり、スマートグラスを用いたものではないため、今後、実機を使用した評価実験を行うことでより実環境を想定した可視化システムの研究を進めていくことが課題である。最終的には実際に聴覚障がい者に使用していただき、可視化システムの性能評価を行いたいと考えている。

参考文献

[1]加藤優, 平賀瑠美, 若月大輔, 松原正樹, 寺澤洋子　“聴覚障害者のための視覚情報を利用した環境音学習の基礎的検討” 情報処理学会研究報告 Vol.2017-AAC-4 No.2 pp.1-5 (2017).

[2]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”　筑波技術大学テクノレポート 24(1), pp.83-84 (2016).

[3]浅井研哉, 綱川隆司, 西田昌史, 西村雅史, “聴覚障害者支援のための環境音可視化システムの開発”情報処理学会研究報告 Vol.2019-AAC-9, No.5, pp.1-8 (2019).

[4]錦織勇人, 西田昌史, 綱川隆司, 西村雅史 ″聴覚障がい者のためのスマートグラスを用いた音声・環境音の可視化システムの構築″情報処理学会第 83 回全国大会講演論文集, 1ZH-01, pp.4_809-4_810 (2021).

[5]平井和之, 西田昌史, 綱川隆司, 西村雅史　“聴覚障がい者を対象としたスマートグラスを用いた日常生活音の認識”情報処理学会第 84 回全国大会講演論文集, 7ZJ-07, pp.4_767-4_768 (2022).

Copyright ©2024 Information Processing Society of Japan.
All Rights Reserved.