

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

錦織勇人 西田昌史 綱川隆司 西村雅史

静岡大学情報学部

1. はじめに

聴覚障がい者は音声によって人とコミュニケーションを取ったり、周囲の状況を把握したりすることが困難である。聴覚障がい者がスポーツ観戦する場合において、競技場の音を認識してスマートグラスに可視化することで聴覚障がい者のスポーツ観戦を支援する研究が行われている[1][2]。他にも人との会話を音声認識してスマートグラスで可視化する研究が行われている[3]。また、会話時の発話印象をスマートグラスで可視化する方法も検討されている[4]。浅井ら[5]は周囲の環境音を認識し、スマートフォンやタブレット等の端末に環境音を可視化することで、聴覚障がい者を支援するという手法を提案している。しかし、スマートフォンやタブレット等の端末を持つために手が塞がることや、端末に目を移して結果を確認しなければならないという問題がある。

人との会話や環境音の可視化は別々のシステムとしてこれまで開発されてきた。しかし、聴覚障がい者にとっては人の声も環境音も理解できる必要がある。そこで本研究では、予め登録されている特定の人の声や環境音に対して、人の声か環境音かを自動で識別して、人の声なら音声認識をして発話内容を可視化し、環境音なら予め登録されている環境音の種類を認識し、スマートグラスで可視化するというシステムを提案する。また、聴覚障がい者がシステムに認識させたい環境音・音声を健聴者のサポートにより登録できる手法を提案する。

2. 提案システム

本研究では音の録音、可視化に用いるスマートグラスとして図1に示すEPSON MOVERIO BT-300を使用する。



図1. EPSON MOVERIO BT-300

本システムはクライアント側（スマートグラス）とサーバー側に分けて設計している。クライアント側では音の録音、結果の可視化を行う。サーバー側では音声か環境音かの判別、音声認識、環境音認識を行う。

Speech and Environmental Sound Visualization Using Smart Glass for Hearing Impaired People
Yuto Nishikori Masafumi Nishida
Takashi Tunakawa Masafumi Nishimura
Faculty of Informatics, Shizuoka University

クライアントとサーバー間でのデータの受け渡しはソケット通信を用いて行う。サーバー側には MacBook Pro (macOS Catalina) を使用する。

本システム内の音声か環境音かを判別する機能と、環境音認識の機能はSVM(Support Vector Machine)を用いている。あらかじめ、スマートグラス付属のマイクを用いて録音した音声、環境音を SVM に学習させるためのデータとして用いている。音声の内容を認識する機能は Julius を用いている。音声と環境音はサンプリング周波数 16kHz で収集し、SVM を用いる際の特徴量にはフレーム長 25ms、シフト幅 10ms で抽出した MFCC24 次元を用いた。また、SVM の認識結果はフレームごとの認識結果の頻度から音イベントの種類を推定している。

本システムの流れを以下の図2に示す。

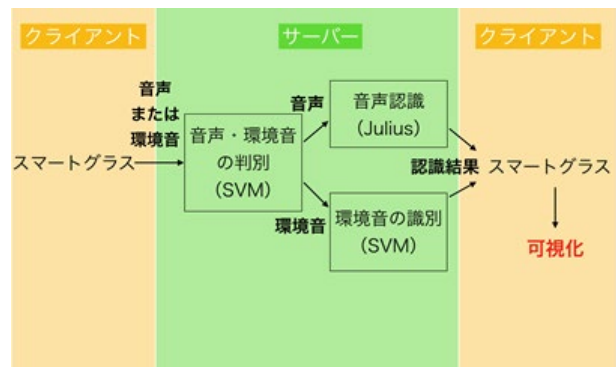


図2. 提案システムの流れ

はじめに、クライアント側で付属のマイクによって周囲の音（環境音または人の声）を録音する。録音した音をサーバー側に送り、送った音を音声か環境音かを SVM によって判別する。判別結果が音声であった場合、Julius で音声認識を行い、認識結果をクライアント側に送り、判別結果が環境音であった場合、どの環境音かを識別するための SVM によって環境音認識を行い、認識結果をクライアント側に送る。クライアント側でスマートグラスによって可視化する。以下の図3にスマートグラスでの認識結果を可視化した画面を示す。



図3. 認識結果の可視化例（左：環境音，右：音声）

さらに、聴覚障がい者が認識させたい音を健聴者に登録してもらうこともできる。登録には音声と環境音を登録することができ、環境音を登録する場合には音が何の音であるかのラベル入力が必要となる。まず、音声を登録するか環境音を登録するかを選択する。環境音を登録する際は登録したい音の名前をテキストボックスに入力し決定ボタンを押す。最後に、録音開始ボタンを押した後、登録する音を発することで SVM が音を学習する。学習終了後、再生ボタンによって登録した音を聞くことができる。すでに学習済みの環境音はラベル入力画面の左側に表示されている。以下の図 4 に実際の目覚ましのアラーム音の登録画面を示す。



図 4. 登録時の画面例(左:ラベル入力,右:音入力)

3. 評価実験

SVM による音声か環境音かの判別、SVM による環境音認識、Julius による音声認識の精度の評価実験を行った。Julius では DNN-HMM によって認識を行った。

評価実験に用いるデータとして、音声データを 10 人(健常者)から予め定めた 25 種類のフレーズの発話を各 3 回ずつ録音した計 750 個、環境音データを 7 種類(体温計の音、携帯の着信音、目覚まし時計のアラーム音、レンジの音、ケトルのボタンの音、インターホンの音、ドアが閉まる音)を 60 回ずつ録音した計 420 個用いた。

音声か環境音かの判別の評価実験には音声と環境音の全てのデータを合わせた計 1170 個を用いた。

本システムでは音声区間の自動検出が未実装であるため、手動で区間を設定したデータを用いた。よって、今回の実験では環境音と音声が続いたり、重なったりするデータを想定していない。

音声・環境音の判別 (SVM) の評価方法を示す。まず、テストデータが音声の場合、環境音のデータは全て学習に用い、音声データは 1 人分(75 個)をテストに用いて、残り 9 人分のデータは学習データに用いた。これを 10 回(10 人分)行い、計 750 個のデータをテストした。次に、テストデータが環境音の場合、音声データは全て学習に用い、環境音データは 70 個をテストに用いて、残り 350 個を学習データに用いた。これを 6 回(6 パターン分)行い、計 420 個のデータをテストした。

環境音の判別 (SVM) の評価方法を示す。環境音データ 420 個を学習に 350 個、テストに 70 個用いた場合に出来る 6 パターンで学習を行い、計 420 個のデータをテスト

した。

音声認識 (Julius) の評価方法を示す。Julius は学習済み音響モデルを搭載しているので、全ての音声データをテストデータとして認識させ、認識結果から文字単位の認識率と発話単位の認識率を出すことで、計 750 個のデータをテストした。

また今回の実験では音声・環境音の判別と環境音認識、音声認識はそれぞれ独立した機能として評価を行なっている。各評価実験の結果を以下の表 1 に示す。

表 1. 実験結果

音声・環境音の判別 (イベント単位)	環境音認識 (イベント単位)	音声認識 (文字単位)
93.5%	98.8%	91.7%

音声・環境音の判別の正解率は 93.5% で高い精度で行うことができた。その中で音声が正しく判別された割合が 92.1%、環境音が正しく判別された割合が 96.0% であった。

環境音認識の認識率は 98.8% で、環境音の識別も高い精度で行うことができた。体温計の音の中でケトルのボタンの音に誤認識されたデータが 2 個、インターホンに誤認識されたデータが 1 個であった。目覚ましのアラーム音の中で体温計に誤認識されたデータが 2 個であった。体温計の音がケトルのボタンの音に誤認識されたのは、どちらも音が短いため無音区間が多くなってしまったことが原因だと考えられる。目覚ましのアラーム音が体温計に誤認識されたのは、どちらもピツという電子音であり音の性質が似ているためだと考えられる。

音声認識の認識率は 91.7% で、これは文字単位の認識率である。発話単位の認識率は 72.5% であった。

4. おわりに

本研究では人の声か環境音かを自動で識別して、人の声なら音声認識をして発話内容を可視化し、環境音なら予め登録されている環境音の種類を認識し、スマートグラスで可視化するというシステムを提案した。また、聴覚障がい者がシステムに認識させたい音声・環境音を健聴者のサポートにより登録できる手法を提案した。

今後の課題としては音声区間の自動検出や音声と環境音以外の雑音の判別などを実装し、より実用的なシステムを構築する必要がある。また、音声・環境音の判別を誤った区間の音声認識率や環境音認識率に対しての影響の考察なども行う必要がある。

参考文献

- [1] 小林雅晃, 西田昌史, 網川隆司, 西村雅史, “聴覚障がい者のための AR クラスを用いたスポーツ観戦支援”, 日本音響学会講演論文集 1-8-4, pp. 1389-1390 (2020).
- [2] 矢島義久, 細井健太, 西田昌史, 網川隆司, 西村雅史, “聴覚障がい者を対象とした競技場の音の認識と可視化”, 情報処理学会第 81 回全国大会 12J-02, pp. 4-775, 4-776 (2019).
- [3] 渡辺大樹, 松本哲也, 竹内義則, 工藤博章, 大西昇, “聴覚障害者のための AR メガネを用いた音声理解”, 情報処理学会研究報告 Vol. 2017-AAC-4, No. 3, pp. 1-6 (2017).
- [4] 北村拓己, 網川隆司, 西田昌史, 西村雅史, “聴覚障がい者のための AR グラスを用いた発話印象の可視化”, 情報処理学会第 82 回全国大会 62D-06, pp. 4-403, 4-404 (2020).
- [5] 浅井研哉, 網川隆司, 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).