

環境音認識を応用した情報提供機能を有するモバイルアプリケーションの検討

中西 恭介 津田 貴彦 西村 竜一 河原 英紀 入野 俊夫

和歌山大学 システム工学部

1 はじめに

iPhone の「Siri」[1] や docomo のスマートフォンで利用できる「しゃべってコンシェル」[2] といった、音声認識機能を持つ音声エージェント機能の実用化が始まっている。これらのシステムでは音声を入力として用いるが、我々は日常生活において、音声だけでなく、周辺の環境音からも多くの情報を得ている。音声からはユーザの意思を理解できるのに対して、環境音からは、その場の状況を判断することができる。

本研究では、環境音からその場の状況を判断し、それに応じて適切な情報を利用者に提示するモバイルアプリケーションの開発を目指す。例えば、グラウンド付近でサッカー部が練習している音を認識することで、そのサッカー部に関する詳しい情報を提示する。

提案システムは、サーバ・クライアント方式によるアーキテクチャを採用しており、Android 端末で録音した音響信号をサーバ側で認識処理する。現在、本研究では、環境音認識の検討 [3] と、音響信号サンプルの収集および Android アプリケーションの実装を行っている。現在までに、環境音を収集するための Android アプリケーションを作成し、それを用いて収集した音響信号サンプルを用いて認識の予備実験を行った。本稿では、その結果を報告する。

2 提案システムの概要

日本国内で普及しているスマートフォンの OS の割合は、Android が 50%以上を占めている。そこで、本研究では、提案システムを Android OS で動作するアプリケーションとして実装している。図 1 に示すように、クライアントプログラムとウェブサーバ上で動作するサーバプログラムの連動によってシステムを構成する。処理の概要は以下ようになる。

1. Android 端末の内蔵マイクロホンで周囲の環境音を収録
2. 収録信号のデータと、収録時間および収録位置情報をインターネットを介して分析用サーバに逐次アップロード
3. 受信データを分析用サーバで統計的認識処理
4. 認識結果をクライアント側に返却
5. 結果に関する情報を Android 端末上で提示

また、サーバ上に収録信号を蓄積することで、認識処理の学習用データとし、利用することを予定している。

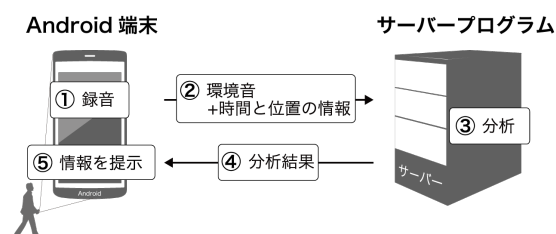


図 1: 提案システムの構成

Android 端末上で動作するクライアントプログラムは Java、サーバ上で動作するプログラムは CGI で実装している。

2.1 検討事項

提案システムは、音響信号等の比較的大容量のデータを扱うため、サーバ・クライアント方式を採用した。このため、提案システムを利用する際にはネットワークに接続している必要がある。しかし、場所によっては、通信環境が常に安定しているとは限らない。また、音響信号の収録時間が長くなると、システムに与える負荷が高くなるだけでなく、情報提供までに要する時間も長くなる問題が一般に生じる。このような状況を想定し、本研究では、収録信号を 1 秒間毎に分割した上で、サーバ上の認識システムの入力とすることを検討した。本研究では、分割された固定長の学習データを大量に用いることで、比較的単純な認識アルゴリズムでも認識性能を確保することを目指す。

3 Android アプリケーションの開発

本研究では、実環境下の音響信号のサンプルを収集するために、環境音収集用アプリケーション「ROCK ON」を作成し、ウェブサイトを通して配布を行っている。サンプリングレート 48kHz、ビットレート 16bit で対象音を録音し、wav 形式で保存するソフトウェアである。ROCK ON には録音以外に、ウェブサーバに収録信号を自動的にアップロードする機能や、GPS による位置情報を取得する機能、および簡単なアノテーション機能を実装している。また、このアプリケーションは、SD カードの容量限界まで連続で録音をすることが可能である。Android 端末を用い、実環境で周囲の雑音も含めて収録することで、利用実態に即した検討を本研究では可能としている。

4 音響信号サンプルの収集

実際に、ROCK ON を用いて、音響信号の収集を行った。収録の条件としては 2 つの場面を想定した。まず、和歌山大学内において、グラウンドや食堂などの各地点で定点で 24 時間録音を行った。収録時の様子を図 2 に示す。付近の植木に Android 端末と充電用のバッテリーを固定し、雨風から保護するために全体をレインカ

Development of Mobile Information System Based on Automatic Environmental Sound Recognition Method, Kyosuke NAKANISHI, Takahiko TSUDA, Ryuichi NISIMURA, Hideki KAWAHARA, Toshio IRINO (Faculty of Systems Engineering, Wakayama University)

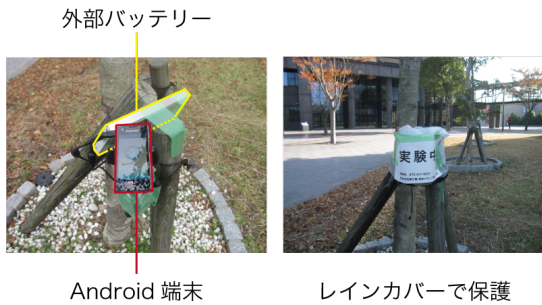


図 2: 音響信号サンプルの収集風景

パーで覆った。グラウンドで録音を行った結果、野球部の練習中の音や体育の授業中の周辺音の収集に成功した。1秒間毎に分割された音響信号データのファイル数は105,070個となった。

次に、今後システムの利用範囲を学外にも広げていくことも想定して、学外の環境音も並行して収集した。この際、個人であらゆる場所の音響信号サンプルを大量に収集することは難しい。このため、楽天リサーチ社のモニタ誘因サービスを利用し、大規模にサンプルを収集することにした。モニタ登録をしている被験者候補に、ROCK ON のダウンロードサイトの URL が書かれたメールを送信した。それを受信した被験者が、所有している Android 端末に ROCK ON をインストールすることで、本実験に参加することができる。環境音を録音し、データをアップロードした被験者には、報酬として会社よりポイントが付与される仕組みである。

様々な地域の環境音を収集するために、モニタの事前スクリーニングを同社に依頼した。スクリーニングでは、Android 端末を所有していることを確認した後、住所や電話番号等の連絡先を記入していただき、居住地が分散するように被験者を選んだ。その結果、日常で耳にする物音やサイレン等の音響信号サンプルを 841 個集めることに成功した。そのうち 588 個は、録音開始時の位置と時間の情報を記録することに成功した。今回、428 人のモニタからデータのアップロードを得ることができた。事前のスクリーニングの結果、863 人のモニタに協力依頼のメールを送信したため、回収率は 50% となった。

本実験によって、GPS の位置情報と時間情報に、周辺の環境音をリンクしたデータベースを作成できるようになった。ただし残りの 253 個のデータに関しては、GPS の位置情報を記録することができなかった。また、アップロードの際にネットワークに何らかのトラブルが発生したため、収録データが途中で欠損しているものもあった。今後更なる改良と、収録データの分析を行う予定である。

5 環境音認識の予備実験

収集した音響信号サンプルの一部を用いて認識の予備実験を行った。今回の予備実験に用いたデータは全部で 10 クラスとした。諸元を表 1 に示す。認識プログラムとしては、GMM(混合正規分布モデル) を音響モデルと

表 1: 実験試料と評価実験結果

クラス	サンプル数	分割数	正解率 [%]
救急車のサイレン	16	374	72
ブザー音	3	21	5
踏切音	24	575	70
ホイッスル	8	20	60
パトカーのサイレン	2	31	0
信号機の電子音	9	161	27
消防車のサイレン	5	193	18
電車のドア	35	95	56
電車内の音	19	242	86
電車外の音	11	148	68
合計	132	1860	-

する音声認識器 Julius 4.2.1[4] を応用した。収録音声 を 1 秒で分割し、一般に、音声認識で用いる MFCC(メル周波数ケプストラム係数) と MFCC、Power を音響特徴量として抽出した。分割前の 1 サンプルを評価データ、残りの全てを認識器の学習用データとし、完全にオープンな状態で交差検定を行った。学習用と評価用を 1 サンプルずつ入れ替えて実験を繰り返した。

実験結果として、表 1 に正解率を示す。「救急車のサイレン」、「踏切音」、「電車内の音」に関しては 70% 以上の高い正解率を得ることができた。一方で、「ブザー」、「パトカー」、「信号機」、「消防車」は、他のクラスに認識される事例がほとんどとなった。これらは、ターゲット信号の継続時間が短く、十分なサンプルが得られなかったことがその原因に挙げられる。

以上より、認識システムの更なる精度の向上が必要であるが、十分なサンプルを集めることで提案システムでの利用は可能であることを確認した。

6 まとめ

本研究では、Android 端末で利用できる環境音収集用アプリケーションを作成し、それを用いて環境音の収集を行った。その結果、学内では 105,070 個、学外では 841 個の音響信号サンプルを収集することに成功した。また、収集した音響信号を用いて予備実験を行い、音声識別器を提案システムに応用することが可能であることを確認した。今後は更なる音響信号の収集を行い、位置と時間の情報を環境音認識アルゴリズムに組み込むことで、認識精度の向上を目指す。また、現在はプロトタイプのレベルである提案システムの実装を進める。

謝辞 本研究の一部は、日本電信電話株式会社との共同研究及び JSPS 科研費 21300066 の助成を受けて行われた。

参考文献

- [1] <http://www.apple.com/jp/ios/siri/>(2013 年 1 月 11 日)
- [2] <http://www.nttdocomo.co.jp/service/information/shabette-concier/>(2013 年 1 月 11 日)
- [3] 津田ら, 携帯型端末で収録した音サインやサイレンなどの環境音認識の検討, 日本音響学会 2012 年秋季研究発表会講演論文集, pp.1515-1516, 2012.
- [4] Lee, A., et al., "Julius - An Open Source Real-Time Large Vocabulary Recognition Engine", Proc. Eurospeech, pp. 1691-1694, 2001.