平成29年度筑波大学情報学群情報科学類卒業研究中間報告

**音響イベント検出における空間情報の活用に関する研究**

ソフトウェアサイエンス主専攻　201411409 溝口和輝

指導教員　山田武志（システム情報系）

牧野昭二（システム情報系）

提出日2017年10月30日

1. 研究背景

　私たちの日常は大量かつ多様な環境音によって囲まれている。環境音に含まれている情報を取り出すことを環境音認識といい、高齢者の見守りシステムや動画の自動タグ付けなどの様々なシステムに対して適用が検討されている。環境音認識には大きく分けて2つの分野がある。足音や車の通過音などの短時間に起こる個々の音響イベントを検出する音響イベント検出と、繁華街やオフィスなどの環境音の録音された状況を識別する音響シーン識別である。しかし、複数の音が同時に存在するような現実の環境下で、これらを確実に実行するのは依然として困難である。

　この問題に対処するため、2017年11月にIEEEのAudio and Acoustic Signal Processing Technical Committee（AASP）が主催の、環境音認識技術を向上させることを目的としたDCASE（Detection and Classification of Acoustic Scenes and Event）2017[1]というワークショップが昨年に引き続き開催される。このワークショップではDCASE2017 Challengeとして音響イベント検出や音響シーン認識を含む4つのタスクが設定されており、参加者は共通のデータセットを用いることで自身の開発した手法の性能の比較評価が行えるようになっている。

1. 研究目的

　本研究では、特に音響イベント検出の性能を改善することを目的とする。この目的を達成するにあたり、ステレオ録音から得られる空間情報を効果的に活用する手法について検討する。昨年開催されたDCASE2016においては、入力としてステレオ録音を用いた手法が総合的に最も良い精度を達成している[2]ものの、空間情報を十分に活用しているとは言い難い。また、DCASE2017において提出された手法のうち、ステレオ録音をモノラル化して利用したものが依然として全体の8割に及ぶ。これらのことから、空間情報をより効果的に活用する方法論を見出すことが急務である。

1. 研究方法

　本研究ではDCASE2017のデータセットを用いた比較評価を行う。提示された4つのタスクのうち本研究で取り組むものは「Sound event detection in real life audio (実生活におけ

る音響イベント検出)」である。このタスクでは図1のように録音されたデータと各録音データに対応したイベントラベルの一覧を元に、未知の音源データからイベントラベルの種類とラベルの開始時間及び終了時間を出力することが求められる。データの録音は実際の道路沿いにて行われ、録音データのラベルにはブレーキ音、車、子供、大型車両、歩行音、会話音の6種類が当てられた。

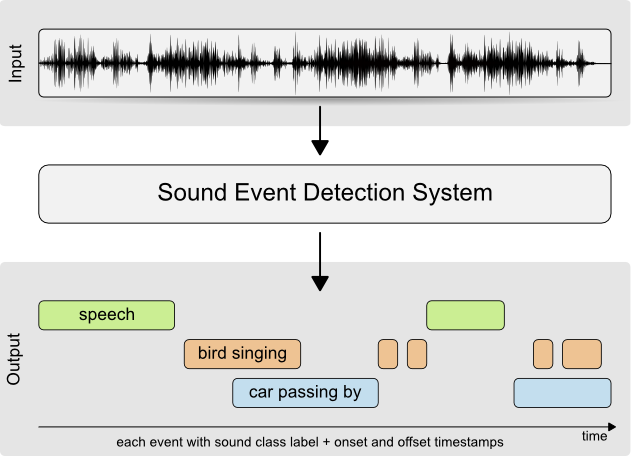


図 1 音響イベント検出の概略

　またDCASE2017ではベースラインとなるシステムが公開されており、参加者はこのベースラインを改良するか、自ら新しいシステムを開発することでベースラインを上回る精度を出すことを目標とする。DCASE2017のベースラインシステムは音響特徴量にLog melband-energyを使用しており、識別器はMultilayer perceptron architecture(MLP)、左右の音を平均化したものを入力に用いたモノラル入力である[3]。このベースラインシステムの精度は42.8%であった。また本タスクで最も高い精度を記録した手法は音響特徴量にScattering transformとClusteringを、識別器にNeuroevolutionを用いた手法であり、精度は44.9%であった[4]。

　音響シーン分類やイベント検出をニューラルネットで行う際に入力として音響特徴量を用いるのが一般的である。しかしLog melband-energyやMel-Frequency Cepstral Coefficients (MFCC)といった音響特徴量は位相情報を持たないのでニューラルネットで位相情報を活用した処理を行うことは出来ない。またDCASE2016の従来手法[2]では入力に左右の音の到着時間差を用いたが到着時間差を使用しなかった場合と比較して精度が9.9%下がっている。よって本研究ではDCASE2017のベースラインシステムに対してフロントエンドで空間情報を抽出しネットワークで利用することで従来手法と比較して精度がどれだけ向上するかを確認する。音響特徴量の次元を2倍にする手法、バーチャルマイクを導入する手法などを導入しそれぞれの手法においてどの程度の効果があるか比較する。

1. 進捗状況

　現在はDCASE2017ベースラインシステムをベースに入力を左右の音を独立して入力出来るように書き換えている。今後の予定としては、現在の実装が終わり次第フロントエンドでの処理を試行する。

1. 参考文献
2. DCASE2017WebSite, http://www.cs.tut.fi/sgn/arg/dcase2017/
3. Sharath Adavanne, Giambattista Parascandolo, Pasi Pertila, Toni Heittola, Tuomas Virtanen,”Sound Event Detection In Multichannel Audio Using Spatial And Harmonic Features,” DCASE2016, 2016
4. https://github.com/TUT-ARG/DCASE2017-baseline-system
5. Christian Kroos and Mark D. Plumbley, ”Neuroevolution For Sound Event Detection In Real Life Audio: A Pilot Study,” DCASE2017,2017