

音源分離を用いた屋外における音響認識システムの開発

金沢工業大学
工学部 情報工学科
中沢研究室
4EP1-60 成瀬 祥馬

研究背景

近年、音の情報を元に、周囲で発生した事象の推定、鳴き声から生物の観測など、音を用いた認識技術に対する様々な取り組みが成されている。
しかし、得られる音には様々な音が混在しており、屋外となればより多数の音が入る。
音の認識をより正確に行うためには、これらの雑音を除去し、必要な音だけを抽出する音源分離が必要である。

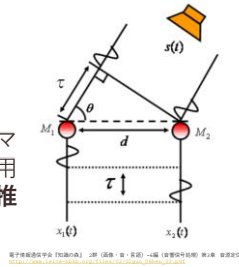
研究目的

得られた音から音源分離技術を用いて必要な音を抽出し、様々な音が混在する屋外において高い精度で周囲の状況を認識できるシステムを開発したい。

音源定位

音の発生した方向を推定する技術

代表的なものとして、音源から複数マイクに対する音の到達の時間差を利用して、音源の方向を推定する**遅延時間推定法**がある



TAMAGO-03



たまご型のマイクロホンアレイ
360° 囲むように45° 間隔で8つのマイクがついており、音源定位に有用である

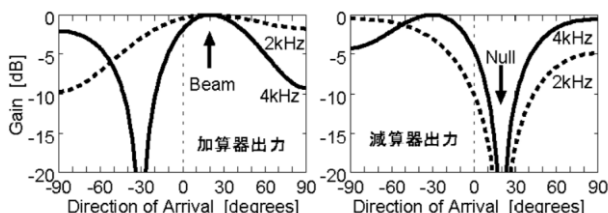
HARKというソフトウェアのサポートデバイスでもあり、HARKを使うことで音源定位、音源分離が比較的容易に行える

音源分離

雑音が含まれる音の中から目的音を抽出する技術。アルゴリズムには様々なものがあるが、代表的なものとして**ビームフォーミング**という手法がある。

ビームフォーミング・・・抽出したい目的音方向を特定(音源定位)し、目的音方向以外の感度を下げることによって、目的音を強調させる事ができる

下図は、複数マイクからの入力を加減算することにより、目的方向の音を強調して表現している。

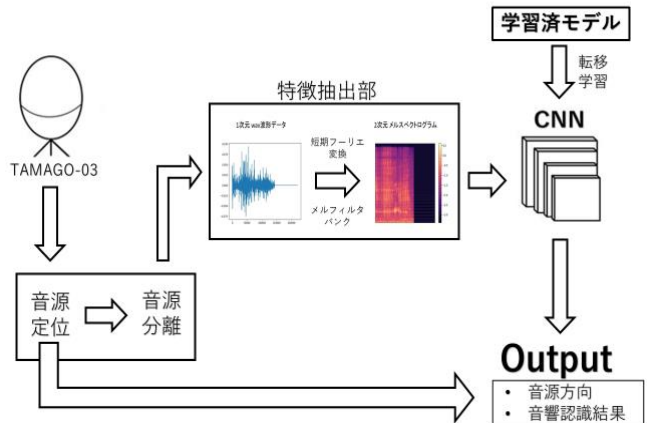


ESC-50データセット

動物	自然音/水の水音	人間の非音声	家庭内音	外部/都市騒音
犬	雨	赤ちゃんの泣き声	ドアノック	ヘリコプター
雄鶏	海の波音	くしゃみ	マウスクリック	チェーンソー
豚	火の弾ける音	拍手	タイピング音	サイレン
牛	コオロギ	呼吸音	ドアのきしみ	クラクション
カエル	鳥のさえずり	咳	缶を開ける音	エンジン音
猫	水滴	足音	洗濯機	列車
妻鳥	風	笑い	掃除機	教会の鐘
昆虫	水を注ぐ音	歯磨き	アラーム	飛行機
羊	トイレの水洗面	いびき	時計の秒針音	花火
カラス	雷雨	飲む	ガラス割れ	ノコギリ

- 学習には、環境音分類に適した5カテゴリ,50クラスのwavデータで構成されたESC-50を用いる
- 本研究では、屋外で聞こえる可能性の低い音は除外し、赤枠で囲まれた34クラスを用いる

システム概要



- 特徴抽出部では、得られた音データにフーリエ変換を行い、周波数特性で表したスペクトログラムにメルフィルタバンクをかけ、人間の聴覚に近い特徴で表したメルスペクトログラムを用いる
- マイクから得られた音を、音源分離によって必要な音以外を除去し、メルスペクトログラムに変換したものをCNNで分類する。
- CNNは、学習済の画像認識モデルを転移学習させたものを用いる
- 音源定位により音の発生した方向は判明している
- どの方向で何が起きたのかをリアルタイムで自動的に判断する

評価方法

- 音識別精度
 - CNNでの分類精度
- 音源定位精度
 - 実際の音源方向との一致率