4Q-04　音源分離された生活音の識別による日常活動の推定

望月 大地　　中島 正晴　　渥美 雅保
創価大学理工学部情報システム工学科

1 はじめに

近年，音声認識を用いたサービスが様々提供されているが，日常空間には音声以外にも色々な生活音が存在する．これら生活音を機械が聞き分ける[1]ためには，それらの定位・分離・識別[2]が必要になるが，HARK[3][4]は，マイクロフォンアレイを用いて複数音の定位・分離を行うソフトウェアである．本研究では，屋内の日常生活において発生しうる生活音から HARK を用いて抽出した MSLS 特徴量を入力として，SVM を用いた多クラス分類により音の種類を識別する．また，「包丁・煮物」音の組み合わせは「料理をしている」など，音の組み合わせからどのような活動がその空間で行われているか把握することを試みる．

2 生活音識別のモデル

図 1 に提案するモデルの構成を示す．本モデルは，Kinect でキャプチャした音の音源定位・分離部，分離された各々の音特徴を用いた音の種類の識別部，音の種類の組み合わせからの活動推定部から構成される．

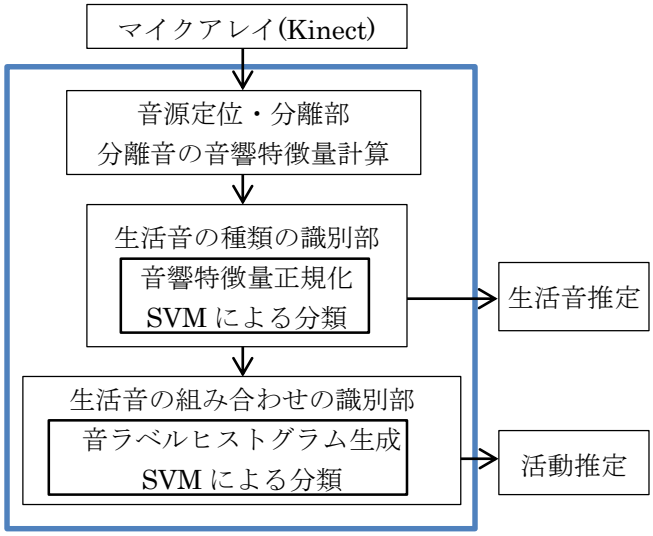

図 1: 生活音識別のモデル

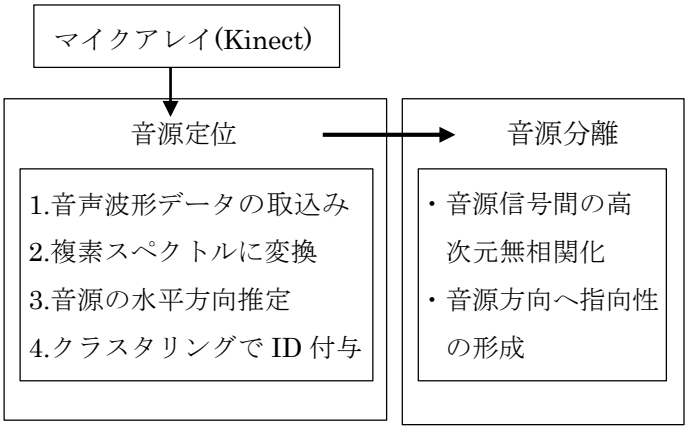

図 2: 音源定位・分離のモジュール構成の概要

3 音源定位・分離

HARK を用いて Kinect でキャプチャした音の音源定位，分離を行う．HARK ではデータフロー指向の GUI 開発環境である FlowDesigner をミドルウェアとして採用している．このことで，音源定位・分離機能を，モジュールを配置・接続していく形式で構築可能である．図 2 に HARK モジュールでの処理の概要を示す．

4 生活音の特徴の学習と識別

サポートベクトルマシン(SVM)を用いた多クラス分類学習により生活音の特徴の学習を行う．SVM の種類はカーネル法を用いた多クラスソフトマージン SVM で，カーネルとしては多項式カーネルを用いる．多項式カーネルは式(1)で与えられる．

$$K(x^{(i)}, x^{(j)}) = \left(\mathrm{gamma} \times x^{(i)} \cdot x^{(j)} + \mathrm{coef0}\right)^{\mathrm{degree}} \quad (1)$$

ここで，$x^{(i)}$，$x^{(j)}$は特徴量ベクトル，gamma，coef0，degree はパラメータである．また，多クラス化にはペアワイズ法を用いる．

前述した音源定位・分離器から，音源ごとにメルスケール対数係数とデルタ対数パワー項を次元要素とする 27 次元の音響特徴量が毎フレーム出力される．学習と識別では，それぞれの音の複数フレームにわたる音響特徴量の系列に対して，それがどの種類の生活音の特徴量であるかのラベル付けをした学習用とテスト用のサンプルデータセットを作成して用いる．また，学習の性能の評価には 5 分割 leave-one-out 交差検証を用いる．識別では，学習したモデルを組み込んだ SVM 分類器を構築し，テスト用のデータセットを用いて生活音の分類を行う．

5 生活音の組み合わせからの活動推定

複数の生活音が同時に聞こえる環境において，それらの音から活動を推定する．そのために，音源分離された各音ストリームの音響特徴を 4 で述べた SVM により識別したラベルの全集合から，ラベルのヒストグラムを作成する．このヒストグラムは，生活音の組み合わせを表現しているので，このヒストグラムをいろいろな音の組み合わせについて多数集めたデータセットを作成して，それらヒストグラムをラベル付けして SVM により学習することにより，音の組み合わせから活動を推定するモデルを作成する．SVM は 4 で用いたのと同様の多項式カーネル多クラスソフトマージン SVM である．

6 実験

6.1 生活音の学習と識別実験

本実験で用いた 16 種類の生活音，及び各々に付与したラベルを表 1 に示す．音響特徴量としては， 27 次元音響特徴量の 10 フレームの列からなる 270 次元の特徴

表 1: 実験で用いた生活音とそのラベル

種類	ラベル	種類	ラベル
赤ちゃん泣き声	1	油で揚げる	9
なべで煮る	2	やかん沸騰	10
固定電話着信音	3	包丁で野菜刻む	11
猫鳴き声	4	ピアノ	12
掃除機	5	シャワー	13
咳き込み	6	いびき	14
犬鳴き声	7	水道水	15
ドライヤー	8	タイピング	16

量を用いた．学習及び識別で用いたサンプルの数は, 16 種類の音についてそれぞれ 6000 個で，合計 96000 個である．学習では，5 分割 leave-one-out 交差検証でモデルを評価した．多項式カーネルのパラメータ gamma, coef0，degree はそれぞれ 1.0，1.0，2.0 である．

この学習モデルを組み込んだ識別器を用いて行った識別結果の混合行列と正確度を表 2 に示す．16 種類の音に関して，正確度の平均は 98.1%で，最も低い「包丁で野菜を刻む」音でも 92.0%と高い分類性能が得られた．これより，用いた音響特徴量とそのサンプル数が適切であったことが確認できた．

6.2 複数音の組み合わせからの活動推定実験

音源分離された複数の音ストリームの音響特徴量から音環境を識別して日常活動を推定するための実験を行った．Kinect のマイクアレイは 4 つであるので，本実験では 2 つの音源から別の音が発生している設定で実験した．本実験で用いた生活音の組み合わせと各々に付与したラベルを表 3 に示す．音響特徴量は 6.1 の実験と同じである．学習及び識別で用いたサンプルの数は，8 種類の生活音の組についてそれぞれ 50 個で, 合計 400 個である．学習では，5 分割 leave-one-out 交差検証でモデルを評価した．多項式カーネルのパラメータは 6.1 の実験と同じである．

この学習モデルを組み込んだ識別器を用いて行った識別結果の混合行列と正確度を表 4 に示す．10 種類の組み合わせに関して正確度の平均は 87.5%であり，ラベル 1

表 3: 生活音の組み合わせ

組合せ	ラベル	組合せ	ラベル
油で揚げる	1	赤ちゃん泣き声	5
水道水		いびき	
なべで煮る	2	咳き込み	6
包丁で野菜刻む		やかん沸騰	
固定電話着信音	3	ドライヤー	7
犬鳴き声		シャワー	
猫鳴き声	4	ピアノ	8
掃除機		タイピング	

表 2: 生活音の識別結果

分類対象ラベル＼分類結果ラベル	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	正確度 (%)
1	2901	11	3	14	0	5	2	0	0	1	40	12	1	3	0	7	96.7
2	0	2972	1	0	4	0	12	0	1	0	1	0	3	0	0	6	99.1
3	0	2	2996	0	1	0	0	0	0	0	0	0	0	1	0	0	99.9
4	10	0	0	2963	1	0	0	0	0	0	19	1	4	2	0	0	98.8
5	0	4	0	0	2989	0	2	0	5	0	0	0	0	0	0	0	99.6
6	0	0	0	0	0	2996	2	0	0	0	0	1	0	0	0	1	99.9
7	1	6	0	0	7	1	2978	0	0	0	1	0	0	0	0	6	99.3
8	0	0	0	0	1	9	0	2957	12	1	0	0	0	1	19	0	98.6
9	6	3	0	0	61	4	1	14	2881	0	5	3	1	0	3	18	96.0
10	7	2	0	7	3	6	8	0	0	2954	1	1	1	5	1	4	98.5
11	40	18	22	33	3	15	7	2	3	4	2759	13	0	4	2	75	92.0
12	0	1	0	0	0	1	2	0	0	0	0	2995	0	0	0	1	99.8
13	3	6	0	17	0	0	0	0	4	0	0	0	2968	1	0	1	98.9
14	9	0	0	4	0	6	0	0	0	0	6	0	0	2975	0	0	99.2
15	0	0	0	0	12	6	0	4	3	0	0	0	0	0	2975	0	99.2
16	0	15	1	1	11	7	1	0	43	0	65	1	3	1	2	2849	95.0

表 4 生活音の組み合わせの識別結果

分類対象ラベル＼分類結果ラベル	1	2	3	4	5	6	7	8	正確度 (%)
1	25	0	0	0	0	0	0	0	100
2	0	25	0	0	0	0	0	0	100
3	0	0	25	0	0	0	0	0	100
4	0	0	0	25	0	0	0	0	100
5	0	0	0	0	25	0	0	0	100
6	0	0	0	0	0	12	13	0	48
7	0	0	0	7	0	4	14	0	56
8	0	0	0	0	0	0	0	25	100

～5, 8 の組み合わせに対してはうまく識別できているが, ラベル 6，7 に誤分類が多いように，特定の音の組み合わせで誤分類率が高くなっている点が検討課題である．

7 むすび

本論文では，生活音の識別，及びそれら生活音の組み合わせから室内での活動を推定するための方法を提案した．そして，実験により個々の生活音を高い正確度で識別できること，及び生活音の組み合わせが分類可能なことを確かめた．今後の課題として，同じカテゴリの生活音でも製品の違い等で音の特徴が異なりうるので，学習データを増やしてより細かいクラス分類を行うなどの工夫をしながら，より多くの音の種類を識別できるように拡張していく．また，それらの音が形成する音環境から様々な日常活動を推定できるように拡張していく．

参考文献

[1]井本桂右，野口賢一，島内末廣，大室仲，羽田 陽一：複数の生活音の出現頻度に基づくユーザ行動の識別手法とコミュニケーションへの応用，画像電子学会 VMA 研究会, Vol.32, pp.1-8, 2012.
[2] 中島正晴：音源定位・分離された音声と生活音の識別に関する基礎的研究, 2015 年度創価大学情報システム工学科卒業論文, 2015.
[3] HARK Document Version 2.1.0. (Revision: 7442)
[4] HARK クックブック Version 2.1.0. (Revision: 7442)