

## 生活音からの部屋種別推定手法の検討

城谷 知葵<sup>†</sup> 光来出 優大<sup>‡</sup> 石田 繁巳<sup>‡</sup> 荒川 豊<sup>‡</sup><sup>†</sup>九州大学工学部電気情報工学科 <sup>‡</sup>九州大学大学院システム情報科学府/研究院

## 1 はじめに

スマートスピーカは近年普及が進んでいるデバイスであり、1台で多くのIoT家電を操作可能である。その際、照明など、同種の機器が複数台ある場合「キッチンの照明をつけて」のように、部屋を指定して対象機器を指定する必要がある。一般に、ユーザは今いる部屋を指定し忘れることが多いため、部屋名が省略された場合にはスマートスピーカは発話者の位置に基づいて操作対象機器を推定する必要がある。

部屋名が省略された場合の対象機器推定手法として、筆者らはスマートスピーカから見てどのような部屋がどちらの方向に存在するかという部屋の間取りを認識した上で、発話者のいる部屋を特定し、対象機器を特定する研究を行っている。これに向けては、部屋の方向及びその方向の部屋の種類を認識することが必要である。

部屋の方向推定については、発話者の方向を推定し、スマートスピーカから見た部屋の方向を推定する仕組みを既に提案している [1]。部屋単位で話者の位置を推定し、その方向を追跡することで部屋が存在する方向を特定できる。

本稿では、スマートスピーカを用いて取得した環境音に基づいて部屋の種類を推定する手法を示す。例えば、スマートスピーカがキッチンに設置されている場合には食器の音や水道の音、寝室の場合には布団が擦れる音やベッドが軋む音など、部屋によって特徴的な環境音が観測される。このような部屋に特有の音をセンシングし、部屋を推定する。本稿では、部屋の方向が既知である前提で、部屋方向の音を分離して取得した後にその環境音から部屋の種類を推定する手法を報告する。

## 2 関連研究

生活音そのものを推定するのではなく、生活音に基づいて部屋の種類を推定する研究は、筆者らの調査し

Estimating the room-type using life sound  
Tomoki Joya<sup>†</sup>, Yudai Mitsukude<sup>‡</sup>, Shigemi Ishida<sup>‡</sup>, and Yutaka Arakawa<sup>‡</sup>

<sup>†</sup>EECS, Kyushu University, Japan

<sup>‡</sup>ISEE, Kyushu University, Japan

<sup>†</sup>joya.tomoki@arakawa-lab.com

<sup>‡</sup>mitsukude.yudai@arakawa-lab.com,

{ishida,arakawa}@ait.kyushu-u.ac.jp

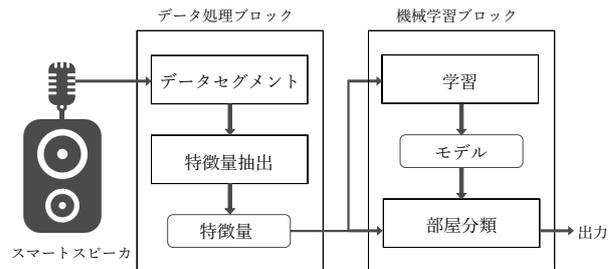


図 1: システム概要図

た範囲では行われていない。音を用いて人間の行動や状況を推定する研究は多く行われている。文献 [2] では生活音を用いて 16 種類の生活音を判別し、それらを総合的に判断することで人間の活動を推定している。また、文献 [3] では、危険を表す環境音や、生活音を判別し、ひとり暮らしの老人を対象とする聞き守りシステムを提案している。これらの手法は本研究においても環境音を推定し、部屋の種類を推定する上で有用である。

## 3 システム構成

図 1 に部屋識別システムの概要図を示す。部屋識別システムは、スマートスピーカ、データ処理ブロック、機械学習ブロックの 3 つで構成されている。スマートスピーカは、使用された直後にマイクを用いて数分間の音データを取得する。スマートスピーカには複数のマイクが搭載されていて方向ごとに音を分離できるものとし、文献 [1] で示す手法で推定した方向の音データを取得する。データ処理ブロックで取得した音データから特徴量を抽出し、機械学習ブロックで学習済みモデルを用いて機械学習により部屋の種類を推定する。以下ではデータ処理ブロック及び機械学習ブロックの詳細を示す。

## 3.1 データ処理ブロック

データ処理ブロックでは、取得した音データをオーバーラップせずに固定長ウィンドウに分割して特徴量を抽出する。特徴量としては、音分析ライブラリの librosa によって算出された、スペクトル中心 (spectral



図 2: 実験環境

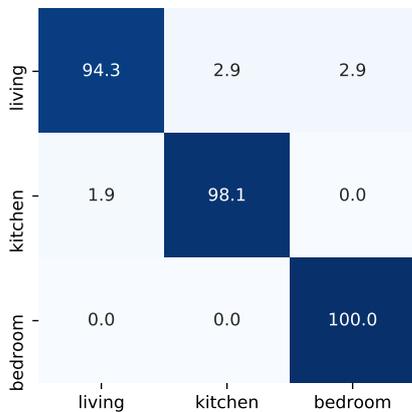


図 3: 混同行列

centroid), スペクトルバンド幅 (spectral bandwidth), スペクトルロールオフ (spectral rolloff), ゼロ交差率 (ZCR), メル周波数ケプストラム係数 (MFCC) を用いる。

### 3.2 機械学習ブロック

機械学習ブロックでは、データ処理ブロックで抽出した特徴量を用いて多クラス分類を行い、部屋の種類を推定する。本研究では機械学習アルゴリズムを限定しないが、本稿では Support Vector Machine (SVM) を使用し、パラメータは  $\text{penalty} = 'l2'$ ,  $\text{loss} = 'hinge'$  とした。

表 1: 評価結果

部屋	適合率	再現率	F 値
リビング	0.97	0.94	0.96
キッチン	0.98	0.98	0.98
寝室	0.97	1.00	0.99
平均	0.97	0.97	0.98

## 4 評価

### 4.1 評価方法

提案手法の有効性を確かめるため、評価実験を行った。図 2 に実験環境を示す。九州大学伊都キャンパス内にある住宅模擬設備内のリビング・キッチン・寝室の各部屋にマイクを設置し、22 歳の男性被験者 1 名がその部屋で起こりうる行動をとった。リビングでは「テレビ」「人の声」、キッチンでは「食器」「皿洗い」「冷蔵庫等の扉の開閉」、寝室では「寝返り」「寝息」の行動を行い、全部屋の扉を開いた状態及び閉じた状態で、行動を行った部屋のマイクを用いてそれぞれ 150 秒間ずつ音データを取得した。サンプリング周波数は 22.05kHz である。データ取得時の機器の操作に関する音を除去するためデータの先頭及び最後をそれぞれ 17 秒ずつを削除した上で、ウィンドウ幅 4 秒間で分割した。各ウィンドウのデータから音声信号処理ライブラリ librosa を用いて特徴量を抽出し、音データを取得したマイクが置かれている部屋がリビング・キッチン・寝室の中のいずれであるかを推定した。評価は 5 分割交差検証により行い、混同行列を用いて適合率、再現率、F 値を算出した。

### 4.2 評価結果

図 3, 表 1 に結果を示す。結果より平均で適合率が 0.97, 再現率が 0.97, F 値が 0.98 と高い精度で判別できている。しかし、リビングの再現率が他の 2 つと比較して低いことがわかる。これは、テレビの音の周波数が多岐に渡り、他の 2 つの部屋と似たような周波数帯であったためと考えられる。

## 5 おわりに

本稿では、部屋ごとの特徴的な生活音に基づく部屋種別推定法を提案した。評価実験を行い、特徴的な音のみを用いることで高い精度で部屋の種類を推定可能であることがわかった。今後は、実際に生活した時のデータを用いた評価実験、間取りの異なる住宅での評価実験などを行う予定である。

## 謝辞

本研究の一部は JSPS 科研費 JP19KT0020 及び東北大学電気通信研究所における共同プロジェクト研究の助成で行われた。

## 参考文献

- [1] 光来出優大, 城谷知葵 他: 2 つのマイクによる部屋レベル話者位置推定の検討, 第 83 回情報処理学会全国大会講演論文集 (2021, 発表予定)
- [2] 望月 他: 音源分離された生活音の識別による日常活動の推定, 第 78 回情報処理学会全国大会講演論文集, Vol. 2016, No. 1, pp. 487-488 (2016).
- [3] 平谷 他: ニューラルネットワークを用いた生活音からの状況認識に関する研究, FIT (第 4 回情報科学技術フォーラム) (2005).