「マルチメディア，分散，協調とモバイル
(DICOMO2021) シンポジウム」 令和3年6月-7月

スマートスピーカを用いた間取り推定手法の初期的評価

石田 繁巳[1,2,a)]　城谷 知葵[3]　光来出 優大[3]　荒川 豊[3]

概要：近年，ネットワークから制御可能なスマート家電が普及しつつあり，音声でスマート家電を操作可能なスマートスピーカの普及も進んでいる．スマートスピーカを用いたスマート家電の操作では，家電の名称に加えてキッチン，リビング，寝室など，家電が設置されている部屋の種類を指定する必要がある．本研究では，部屋の種類を省略した場合の家電操作に向け，部屋の間取りを推定した上で発話者がいる部屋を認識する手法を提案する．今後のスマートスピーカには発話者の方向を取得するためにマイクロフォンアレイが搭載されると想定し，マイクロフォンアレイを用いて取得した音の到来方向を解析することでキッチン，リビング，寝室など，どのような種類の部屋がどちらの方向に存在するかという「間取り」を推定する．その上で発話者の方向を推定し，間取り推定結果と照らし合わせることで発話者がいる部屋の機器を操作対象とする．本稿ではこの実現に向けた第 1 歩として，間取り推定手法について報告する．1LDK の住宅模擬環境内で取得した音データを用いて初期的評価を行った結果，3 つの部屋について部屋の方向を正解率 0.850 で，部屋の種類を正解率 0.474 で推定できることを確認した．

1. はじめに

無線通信技術，IoT (Internet of Things) 技術などの発展とともに，ネットワークから制御可能なスマート家電の普及が進んでいる．ネットワーク家電はこれまでの家電と同様にリモートコントローラを用いて操作を行うことができるほか，スマートスピーカなどによる音声制御を用いた操作も一般的に行われている．スマートスピーカは対話型の音声操作機能を提供するスピーカであり，スマートスピーカ自身の操作のほかにも電気やテレビの ON・OFF の他，テレビのチャンネル操作や音量操作など，ネットワーク家電の様々な操作をユーザが声をかけるだけで実現できる．

スマートスピーカを用いた家電の操作では，例えば「電気を点けて」などのように「どの家電」に対して「どのような操作」をするかという 2 つの指示を与える必要がある．電気やテレビなど，家電の中には複数の部屋に設置され得るものも存在する．複数の部屋に設置された家電を操作する場合には，その家電が設置されている部屋の名称などの場所情報をあらかじめスマートスピーカに設定した上で場所も含めて操作対象の家電を指定する必要がある．

しかしながら，部屋の名称などの場所情報は複数の部屋に設置されている家電を操作する際であっても省略される場合がある．ユーザは自分が今滞在している部屋にある家電を操作することが念頭にあるため，同一の部屋にある家電を操作する場合に部屋の指定を忘れがちである．また，「浴室」という部屋の名称をスマートスピーカに登録しているにも関わらず「風呂の電気を消して」など別の名称で部屋を指定してしまうなど，指示内容に曖昧性が含まれる場合も発生する．このような曖昧な指示に対してはユーザのコンテキストに基づいて操作や対象家電を推定する手法が報告されているが [1,2]，その実現に向けてはユーザ位置などのセンシング情報や過去の操作履歴などに基づく機械学習が必須である．

このような問題に対し，本研究ではスマートスピーカを用いて自動的に「間取り」を認識することで操作対象家電を特定する手法を提案する．まず，スマートスピーカを設置したらキッチン，リビングなど，どのような種類の部屋がスマートスピーカから見てどちらの方向に存在するかという「間取り」を推定する．ユーザがスマートスピーカに対して音声コマンドを発声するとユーザがいる部屋の種類を特定し，ユーザがいる部屋にある機器を操作対象とする．

本稿では，この実現に向けた初期的検討として，スマートスピーカを用いた間取り推定手法を示す．現在市販されているスマートスピーカの多くは単一のマイクロフォンを搭載しているが，発話者の位置推定に向けて今後のスマートスピーカにはマイクロフォンアレイが搭載されると想定される．そこで，マイクロフォンアレイを用いて取得した

[1] 公立はこだて未来大学システム情報科学部
Sch. Systems Information Science, Future Univ. Hakodate
[2] 九州大学システム LSI 研究センター
SLRC, Kyshu University
[3] 九州大学大学院システム情報科学府/研究院
ISEE, Kyushu University
[a)] ish@fun.ac.jp

音の到来方向を分析することでどちらの方向に部屋が存在するのかを特定する．その上で，部屋方向から生じる生活環境音を分析し，水道の音や食器の音，テレビの音など，一般的なキッチンやリビングで発生する音を用いて構築した学習済みモデルに基づいて機械学習により部屋の種類を推定する．

提案の実現可能性を検証するため，九州大学の 1LDK 住宅模擬環境にスマートスピーカに見立てたマイクロフォンアレイを設置して生活行動音データを収集し，初期的な評価を行った．その結果，3 つの部屋について部屋の方向を正解率 0.850 で，部屋の種類を正解率 0.474 で推定できることを確認した．

本稿の構成は以下の通りである．2 では音声の到来方向推定に関する関連研究を示し，3 で提案する間取り推定手法を示す．4 で初期的評価を行い，最後に 5 でまとめとする．

2. 関連研究

筆者らの調査した範囲では，マイクロフォンアレイを用いて部屋の方向を推定する研究はこれまでのところ報告されていない．

マイクロフォンアレイを用いて音源の方向を推定する技術は音源位置推定技術と呼ばれ，広く研究が行われている．代表的な音源位置推定手法として，マイクロフォンで受信した信号のスペクトル分析による手法 [3, 4] や複数のマイクロフォンで受信した信号の到来時間差や周波数成分の差の分析による手法 [5–9]，回折情報と周波数特性の分析による手法 [10] などが報告されている．

受信信号のスペクトル分析による手法 [3] では，複数のセンサを室内に広く分布するように設置し，室内の音源から生じる音波を球面波として仮定して受信した信号のスペクトルを分析することで音源の位置を推定している．しかし，この手法では十分な推定精度を得るために多数のセンサを特定の形状になるように室内に広く分布させる必要がある．

複数のマイクロフォンで受信した信号の到来時間差や周波数成分の差の分析による手法は数多く報告されている．遅延和法 [11] では，空間の各方向に対して各チャンネルの受信信号が同位相になるように調整し，調整した信号を加算することで空間的なパワー分布を求める．ビームフォーミング法 [12] では，複数のマイクロフォンを空間的に異なる位置に配置し音源からの信号を受信し，受信した信号間の到来時間差から信号の届いた方向を推定することで音源の位置を推定している．しかしこれらの手法では多数のマイクロフォンを用いて音源位置の推定精度を高めているため必要な機器の数や装置のサイズが大きくなってしまうといった問題がある．相互相関法 [13] では複数チャンネルの受信信号間の伝達時間差を用いて音源の位置を推定する．チャンネル間の受信信号をずらしながら相互相関係数を計算し，相互相関係数が最大となる時にずらしたサンプル数から伝達時間差を求め音源の位置を推定している．この手法は伝達時間差を用いる他の手法に比べ計算が単純で処理も速いといった特徴があるが，屋内環境など反射音が多い環境では推定誤差が大きくなってしまうといった問題もある．MUSIC (Multiple Signal Classification) 法 [9] では受信した複数チャンネルの信号から空間相関行列を求め，音源成分の固有ベクトルが張る空間と雑音成分の張る空間の直交性を用いて音源の位置を推定している．この手法は他の手法に比べ少ないマイクロフォン数でも高い分解能で音源の位置推定が可能である．

回折情報と周波数特性の分析による音源位置推定手法 [10] では，球状の剛体に複数のマイクロフォンを埋め込み，仮定した音源方向の回折情報から求めたスペクトルと実際のスペクトルとの差が最小になるような方向を実際の音源の方向として推定している．

いずれの音源位置推定手法においても反響や残響の多い屋内環境でこれらの手法を用いる場合には，反射音の影響を軽減することが課題の 1 つであることから，反射音の影響を軽減する研究も報告されている．鈴木らは，マイクロフォンに先行して届く直接音の振幅を保持して後続の反射音をマスクするピークホールド処理を帯域毎に適用し，反射音の影響を低減する手法を報告している [14]．Okamoto らは，マイクロフォンアレイを複数のサブアレイに分割しサブアレイ毎の空間行列を平均化して反射音の影響を軽減する空間平均化法を 3 次元空間に適用し，反射音に頑健な音源位置推定手法を報告している [15]．

Ishi らは，天井に取り付けた 16 個のマイクロフォンアレイと空間の情報を用いて反射音の影響を軽減しつつ複数音源の位置を推定する手法を報告している [16]．この手法では，マイクロフォンに届く複数の反射音の到来方向を推定し，あらかじめ学習しておいた空間の 3 次元地図を用いてそれらの推定結果を統合することで，反射の多い屋内環境での音源位置推定を実現している．また，Ribeiro らは部屋の 3D モデルを構築しそのモデルを用いて音源からマイクロフォンに届く初期反射音を予測して反射を補正することで反射音に頑強な音源位置推定手法を報告している [17]．

このように反射の多い屋内環境で反射音の影響を軽減しつつ音源の位置を推定する手法は多く存在する．しかし，これらの手法では必要とするマイクロフォンの数が 16 個などと多くなることや複雑な信号処理が必要であることが課題としてあげられる．また屋内環境や部屋の空間情報を正確にモデリングした高度な情報を作成するには時間がかかり複雑な処理も必要になるといった課題もある．そのため，これらの手法を用いてスマートスピーカによる音源の位置推定を行うには多くの課題が残り，スマートスピーカに搭載されている少ないマイクロフォン数で反射音の影響を軽減しつつ高い精度で方向推定を実現することは現実的

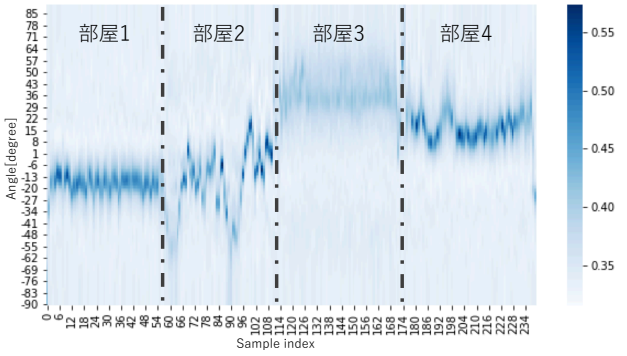


図 1　1 つの移動音源が存在するときの Sound Density Map の例

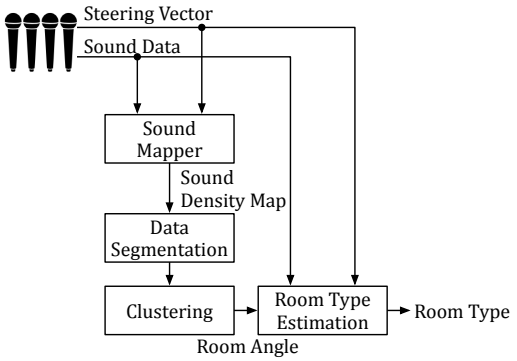


図 2　スマートスピーカを用いた間取り推定手法の概要

ではない．

筆者らはこれまでに，1 人の発話者を対象として，2 台のマイクロフォンを用いて話者の方向の部屋を推定する手法 [18] 及び部屋の種類を推定する手法 [19] を報告した．しかしながら，これらの手法は発話者が 1 人であることを前提としており，複数人が生活する環境に適用することが困難である．

3. スマートスピーカを用いた間取り推定手法

3.1 アプローチ

スマートスピーカを用いた間取り推定の基本アプローチは，「Sound Density Map」を用いて反響の特徴を抽出し，音源を部屋ごとに分離することである．Sound Density Map は音の到来方向の時間変化を描いたマップである．屋内環境で発生した音をマイクロフォンで観測する場合，壁や天井などでの反響の影響を受けて音源の方向推定結果にはゆらぎが生じるため，Sound Density Map 上には各時刻で幅を持った「帯」が現れる．この帯の様相は，部屋の大きさや部屋に置いてある物及びその位置，壁・天井の素材などの影響によって変化することから，Sound Density Map 上の帯の特徴を抽出し，教師なし学習により部屋を分離する．

図 1 は，移動音源，例えば掃除機が 1 つある場合の Sound Density Map を描いた例を示している．1LDK の住宅模擬環境の 1 室にマイクロフォンアレイを設置し，音源を移動させながら取得した音声データを用いて MUSIC 法により到来方向を推定することで描いた．図 1 において移動音源は破線で示した時刻に別の部屋の移動しており，部屋ごとに帯の幅やゆらぎ方が変化していることが分かる．

現実環境では複数の音源が存在することから，Sound Density Map 上には複数の音源に対応する複数の帯が現れる．このため，Sound Density Map 上で音源を分離した上で，到来方向のゆらぎ方などの特徴量を抽出して教師あり学習により部屋を分離する．

3.2 想定環境

本稿で提案する間取り推定手法は，複数の部屋が同じ階層にあり，扉を介して隣接している 2LDK などの住宅環境での利用を想定している．その中の 1 室にマイクロフォンアレイを有するスマートスピーカが設置されており，スマートスピーカを設置した部屋及びその部屋と扉を介して隣接する部屋が間取り推定の対象である．屋内では複数人が活動しており，音は同時に複数箇所で発生することを前提とする．推定対象となる部屋の数のみが既知であり，スマートスピーカからみた各部屋の方向は分からないものとする．

3.3 設計概要

図 2 に，スマートスピーカを用いた間取り推定手法の概要を示す．提案する間取り推定手法は Sound Density Map 描画ブロック，データ分割ブロック，クラスタリングブロック，部屋種別推定ブロックの 4 つのブロックで構成される．まず，スマートスピーカに搭載されたマイクロフォンアレイで取得した音データを用いて，Sound Density Map 描画ブロックにおいて音の到来方向情報の時間変化を描いた Sound Density Map を描画する．データ分割ブロックでは Sound Density Map 上の点を音源ごとに分割し，クラスタリングブロックで各音源を部屋ごとにグループして部屋の方向を推定する．最後に，得られた部屋方向の音を合成し，教師あり学習によって部屋の種類を推定する．

以降では各ブロックの動作について詳述する．

3.4 Sound Density Map 描画ブロック

Sound Density Map 描画ブロックでは，MUSIC 法を用いて音の到来方向情報 $\overline{P_{\mathrm{MUSIC}}}$ を求め，その時間変化を Sound Density Map として描く．MUSIC 法は音源到来方向推定手法の中でも分解能が高いことが特徴の 1 つであり，到来方向のゆらぎを分析する提案手法で特に有効である．

図 3 に Sound Density Map 描画ブロックの概要を示す．図 3 に示すように，Sound Density Map 描画ブロックには音声データに加えてマイクロフォンアレイの物理的な配置

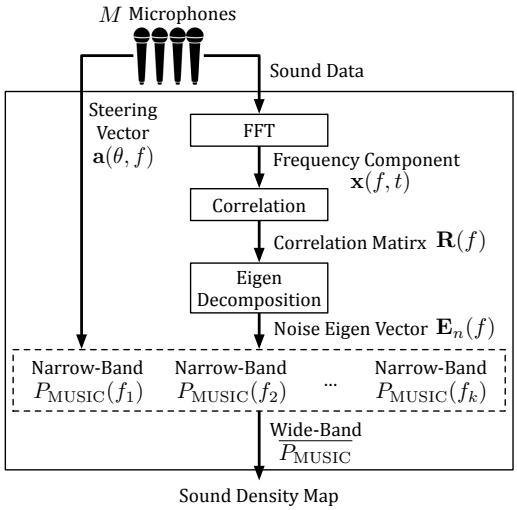


図 3 Sound Density Map 描画ブロックの概要

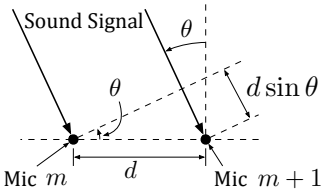


図 4 到来方向とマイクロフォンごとの位相差

情報から計算したステアリングベクトルを入力する．

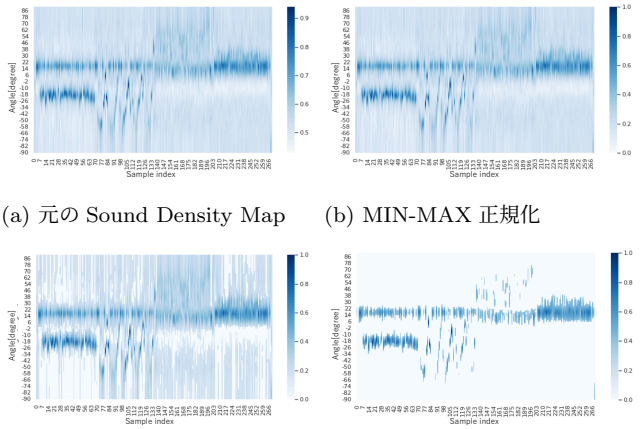
まず，音声データを固定幅のウィンドウで区切り，FFT (Fast Fourier Transform) を施し，周波数成分 $\mathbf{x}(f, t)$ に変換する．ここで，f は周波数，t は時刻である．マイクロフォンの数を M とすると $\mathbf{x}(f, t)$ は M 次元の縦ベクトルであり，各マイクロフォンで取得した音声データを FFT した結果を並べたものである．

次に，求めた周波数成分 $\mathbf{x}(f, t)$ から次式で相関行列 $\mathbf{R}(f)$ を求める．

$$\mathbf{R}(f) = E[\mathbf{x}(f, t)\, \mathbf{x}^H(f, t)] \tag{1}$$

ここで，$\mathbf{z}^H$ はベクトル $\mathbf{z}$ のエルミート転置を，$E[\,]$ は平均化処理を表している．そして，相関行列 $\mathbf{R}(f)$ の固有値を計算し，複数のウィンドウで計算した固有値の大きさの分布から信号成分と雑音成分の数を特定する．信号成分が $N(< M)$ 個であると特定できたとすると，残りの $M - N$ 個の固有値に対応する固有ベクトルを並べて雑音固有ベクトル $\mathbf{E}_n(f)$ を得る．

ステアリングベクトルは，マイクロフォンアレイの物理的配置情報から算出する．ステアリングベクトルは到来方向ごとに各マイクロフォン間で音声データにどの程度の位相差が生じるかを示すベクトルである．**図 4** に示すように M 台のマイクロフォンが等間隔 d で 1 直線に並んでいる場合，角度 θ の無限遠から到来した音はマイクロフォンに距離差 $d \sin\theta$ で到達する．この距離は音速を c として位相差 $2\pi f d \sin\theta / c$ に相当することから，$\phi = 2\pi f d \sin\theta / c$ と



(a) 元の Sound Density Map　(b) MIN-MAX 正規化

(c) 各時刻でピークから −6 dB ま　(d) 各時刻で上位 10%以内を抽出
でを抽出

図 5 Sound Density Map に対するフィルタリング処理の概要

置くとステアリングベクトルは以下のように表される．

$$\mathbf{a}(\theta, f) = \begin{bmatrix} 1 & e^{-j\phi} & e^{-j2\phi} & \cdots & e^{-j(M-1)\phi} \end{bmatrix}^T \tag{2}$$

ここで，$\mathbf{y}^T$ はベクトル $\mathbf{y}$ の転置を表している．

狭帯域の到来方向情報 P_{MUSIC} は，各周波数成分で求めた雑音固有ベクトル $\mathbf{E}_n(f)$ と各周波数のステアリングベクトル $\mathbf{a}(\theta, f)$ とから以下のように求まる．

$$P_{\mathrm{MUSIC}}(\theta, f) = \frac{1}{\mathbf{a}^H(\theta, f)\, \mathbf{E}_n(f)\, \mathbf{E}_n{}^H(f)\, \mathbf{a}(\theta, f)} \tag{3}$$

最後に，全ての周波数成分について狭帯域の到来方向情報 P_{MUSIC} を平均することで広帯域の到来方向情報 $\overline{P_{\mathrm{MUSIC}}}$ を時刻ごとに求める．周波数は離散値であるから，k 個の周波数成分があるとすると，

$$\overline{P_{\mathrm{MUSIC}}} = \frac{1}{k} \sum_f P_{\mathrm{MUSIC}}(\theta, f) \tag{4}$$

となる．広帯域の到来方向情報 $\overline{P_{\mathrm{MUSIC}}}$ は角度 θ 及び時刻 t の関数である．角度 θ を掃引すると，$\overline{P_{\mathrm{MUSIC}}}$ は音源のある方向でピークを示すことから，$\overline{P_{\mathrm{MUSIC}}}$ のピークを検出することで音の到来方向を推定できる．

本提案では，ピークを検出した結果ではなく $\overline{P_{\mathrm{MUSIC}}}$ が時刻とともにどのように変化しているのかを「Sound Density Map」として描く．音が反響している場合には，複数の方向から同じ音が到来するために $\overline{P_{\mathrm{MUSIC}}}$ のピークの尖度が低下するなどの影響が生じる．このため，これらの影響を消すことなく Sound Density Map を描き，推定ブロックにおいて利用する．

Sound Density Map には間取り推定に必要となる音以外にも環境ノイズの影響などによる音到来方向情報が含まれている．間取り推定に必要な音以外の影響を削減するため Sound Density Map に対してフィルタリング処理を行う．**図 5** に，このフィルタリング処理の概要を示す．まず，Sound Density Map における全時刻の $\overline{P_{\mathrm{MUSIC}}}$ に

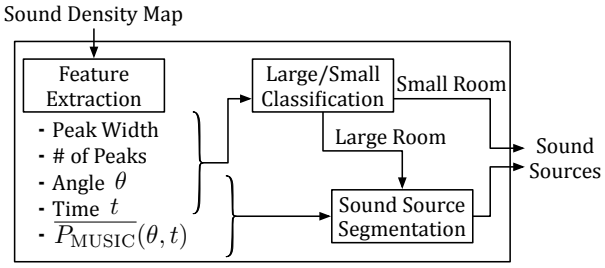


図 6　データ分割ブロックの概要

対して MIN-MAX 正規化処理を行い (図 5b)，時刻ごとに $\overline{P_{\mathrm{MUSIC}}}$ の複数のピークを検出して各ピークから −6 dB までの部分のみを抽出する (図 5c)．各ピークから −6 dB までを抽出することで，各音源に関するエネルギー成分のほとんどを抽出し，それ以外のノイズ成分を除去する．最後に，時刻ごとに $\overline{P_{\mathrm{MUSIC}}}$ の値が上位 10%の部分のみを抽出する (図 5d)．これにより，環境音などによって生じたピークを除去する．

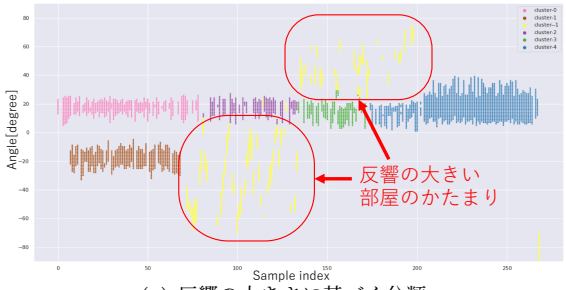
3.5　データ分割ブロック

データ分割ブロックでは，Sound Density Map 上の各点を音源ごとに分割する．1 つの音源の場所は短時間に大きく変化しないため，同一音源からの音は Sound Density Map 上で 1 つの帯として現れる．このため，複数の音源が存在する場合には Sound Density Map 上に複数の「帯」が現れる．そこで，密度に基づくクラスタリング手法である DBSCAN を Sound Density Map に対して適用し，帯を分離することで音源ごとにデータを分割する．

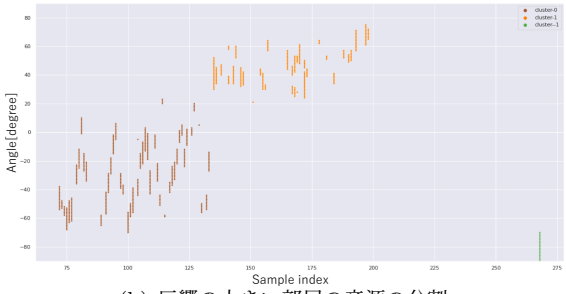
図 6 にデータ分割ブロックの概要を示す．まず，Sound Density Map の各時刻における $\overline{P_{\mathrm{MUSIC}}}(\theta, t)$ のピーク幅，ピーク数を抽出する．そして，これら 2 つの値と角度 θ，時刻 t を特徴量として DBSCAN によるクラスタリングを行う．各クラスタは 1 つの音源に対応する Sound Density Map 上の点群となる．

次に，分割された各クラスタ，すなわち各音源は反響の大小によって 2 つに分類する．各クラスタについて角度 θ の分散を算出し，角度の分散が大きいクラスタと小さいクラスタに分類する．角度分散が大きい部屋は反響が大きい部屋 (あるいはあらゆる方向に音源が存在し得る部屋) の音源であり，分散が小さい部屋は反響が小さい部屋 (あるいは隣接する部屋からドアを介して音が伝わる場合など，音のする方向が限定されている部屋) の音源である．

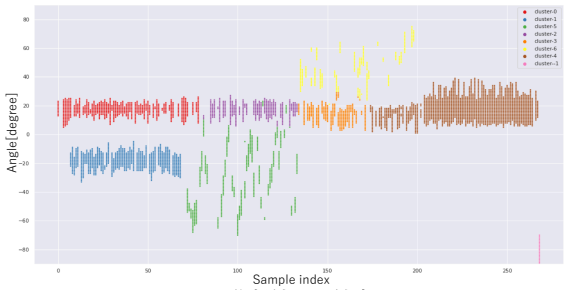
反響の大きい部屋の音源は角度に関する情報だけでは音源を分離できていない可能性があるため，さらに分割を試みる．反響の大きい部屋の音源に分類された Sound Density Map 上の各点において，角度 θ，時刻 t，$\overline{P_{\mathrm{MUSIC}}}(\theta, t)$ という 3 次元の情報を抽出し，DBSCAN を用いて再びクラスタリングを行う．



(a) 反響の大きさに基づく分類



(b) 反響の大きい部屋の音源の分割



(c) 分割結果を統合

図 7　データ分割ブロックの処理概要

最後に，全ての分割結果を統合して音源に基づくデータ分割が完了する．

図 7 は実際のデータ分割の様子を示している．角度の分散の大きさに基づいて反響の大小でクラスタを分類し (図 7a)，反響の大きい部屋はさらに分割する (図 7b)．図 7c は全ての分割が完了した状態を示している．本節で示したデータ分割では同一の部屋の音源を分割できない場合があるが，本提案の目的に鑑みて，同一の部屋に存在する音源は分離する必要はない．

3.6　クラスタリングブロック

クラスタリングブロックでは，データ分割ブロックで分離した各音源を部屋ごとに分割する．DBSCAN によるデータ分割では，音源ごとにデータを分割することを目的としるため，一般には部屋数よりも多くのクラスタに分割される．このため，音源ごとに部屋に関する特徴量を抽出し，同一の部屋で発生した音をグループ化する．

図 8 にクラスタリングブロックの概要を示す．まず，データ分割ブロックで分割した各音源に対応する Sound Density Map 上の点群からクラスタリングに用いる以下の特徴量を抽出する．これらの特徴量は音源が存在する部屋

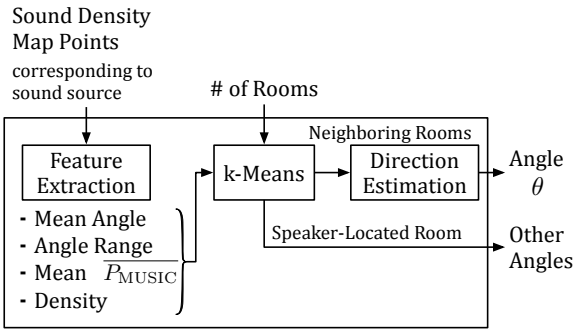


図 8 クラスタリングブロックの概要

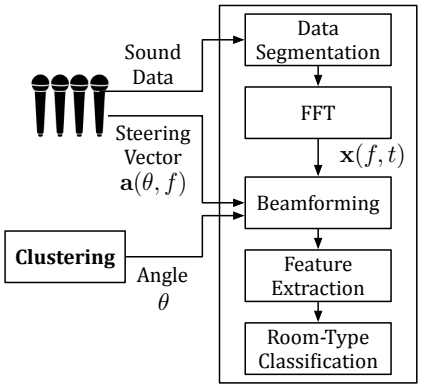


図 9 部屋種別推定ブロックの概要

の方向や部屋の大きさ，反響の仕方などの影響を受ける．

(1) 角度の平均: クラスタ内の全点の角度を平均した値である．

(2) 角度の幅: クラスタ内の全点の角度の最大値と最小値の差である．

(3) 平均到来方向情報: クラスタ内の全点の $\overline{P_{\mathrm{MUSIC}}}$ を平均した値である．

(4) 密度: クラスタの点群が存在する矩形領域の中で，クラスタ点群が存在する割合である．

次に，特徴量を用いてクラスタリング手法により音源を部屋ごとに分割する．クラスタリングには k-means 法を用いる．3.2 で述べたように，想定環境においては部屋の数は既知であるものとし，k-means のクラスタ数は部屋の数，すなわちスマートスピーカを設置した部屋及びその部屋と扉を介して隣接する部屋の合計数とする．

分割されたクラスタには，スマートスピーカを設置した部屋とその部屋と扉を介して隣接する部屋とが含まれる．隣接した部屋で発生した音は，ドアを介してスマートスピーカに到達するため，音の発生する方向はドアの方向に制限される．そこで，クラスタ内の角度の最頻値を求めることで部屋の方向を推定する．

一方で，スマートスピーカを設置した部屋では障害物がなければあらゆる方向から音が発生するため，部屋の方向を定義することができない．このため，角度の分散に基づいてスマートスピーカを設置した部屋に対応するクラスタを特定し，部屋方向推定処理から除外する．

3.7 部屋種別推定ブロック

部屋種別推定ブロックでは，MUSIC ブロックで使用したステアリングベクトルを用いて部屋方向の音を合成し，その音を分析して部屋の種類を教師あり学習により推定する．**図 9** に部屋種別推定ブロックの概要を示す．

まず，音の二乗平均平方根 (RMS) が十分に大きい区間を抽出した上で，固定幅のデータに分割する．取得する音声データには音が無い区間も含まれると考えられるため，RMS が大きい区間のみを取り出すことで有意な音のみを分析対象とする．

次に，分割した各データに FFT を施して周波数成分の時間変化 $\mathbf{x}(f,t)$ を計算する．そして，算出した周波数成分 $\mathbf{x}(f,t)$ と MUSIC ブロックで算出したステアリングベクトル $\mathbf{a}(\theta,f)$ とから部屋方向の音声を合成する．3.4 と同様に，M 台のマイクロフォンが受信した音声データの周波数成分を $\mathbf{x}(f,t)$ とすると，角度 θ 方向の音 $y(f,t)$ は以下のように合成できる．

$$y(f,t) = \mathbf{a}^T(\theta,f)\,\mathbf{x}(f,t) \tag{5}$$

ここで，$\mathbf{a}(\theta,f)$ は式 (2) で示したステアリングベクトルである．$\mathbf{a}(\theta,f)$，$\mathbf{x}(f,t)$ はともに M 次元の縦ベクトルであることから，$y(f,t)$ はスカラー値となる．

合成した $y(f,t)$ を固定時間幅ウィンドウで区切り，部屋種別推定に必要な特徴量を抽出する．特徴量は，マイクを用いた生活行動認識に関する文献 [20] を参考にして以下の 3 種類を用いる．

(1) MFCC の基本統計量: 各ウィンドウで 20 個のメル周波数ケプストラム係数 (MFCC) を計算し，各 MFCC について基本統計量を算出する．

(2) ゼロ交差率の基本統計量: 音声データの時間領域波形において，振幅の正負が入れ替わった割合である．各ウィンドウでゼロ交差率を計算し，その基本統計量を算出する．

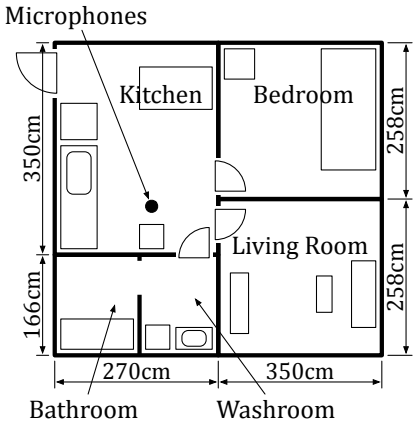
(3) 二乗平均平方根 (RMS) の基本統計量: 各ウィンドウで RMS を計算し，その基本統計量を算出する．

基本統計量は，平均，最大値，最小値，分散，尖度，歪度の 6 種類を求め，132 次元の特徴量を抽出する．

抽出した特徴量を用いて，教師あり学習を用いた分類器により部屋種別を推定する．学習モデルは一般的な住宅環境で収集した音声データを用いてあらかじめ構築しておく．使用する機械学習アルゴリズムは限定しないが，本稿では Random Forest を用いる．

4. 評価

提案する間取り推定手法の実現可能性を確認するため，九州大学伊都キャンパス内の 1LDK 住宅模擬環境で取得し



(a) 実験環境の間取り

(b) 実験の様子

図 10 実験環境

た音声データを用いて初期的評価を行った．本稿で示した間取り推定は，大きく分けると部屋方向の推定，部屋の種別推定の 2 つのタスクを実施していることから，初期的評価ではこれら 2 つのタスクそれぞれの性能を検証した．

4.1 評価環境

図 10 に実験環境を示す．図 10a は実験環境となっている 1LDK 住宅模擬環境の間取りを，図 10b は実験の様子を示している．4 台の AZDEN SGM-990 マイクロフォンを図 10a に示すように壁から 1 m，高さ 70 cm で三脚を用いて 5 cm 間隔で設置し，Behringer UMC404HD USB オーディオを用いてサンプリング周波数 44.1 kHz，量子化ビット数 16 bit で音データを収集した．なお，本稿で示す評価に用いたデータの取得実験は，九州大学倫理審査委員会の承認 (承認番号: シス情認 2020-05) を得て実施した．

4.2 部屋方向推定性能

部屋方向推定の性能を検証するため，2 人の被験者が声を出しながら部屋の中を自由に移動する間に取得した音データを用いて評価を行った．評価には**表 1** に示すデータセットを用いた．各データセットは 40 秒間の録音 20 個で構成されている．各録音は，被験者の 1 人が 1 つの部屋内を自由に移動，もう 1 人の被験者が表 1 に書かれた順に 10

表 1 部屋方向推定性能用のデータセット

Dataset (40 s × 20)	Sound Source 1 (Subject A Voice)	Sound Source 2 (Subject B Voice)
Bedroom DS	Bedroom	Bedroom (10 s)
Kitchen DS	Kitchen	→ Kitchen (10 s)
Washroom DS	Washroom	→ Washroom (10 s)
Living DS	Living Room	→ Living (10 s)

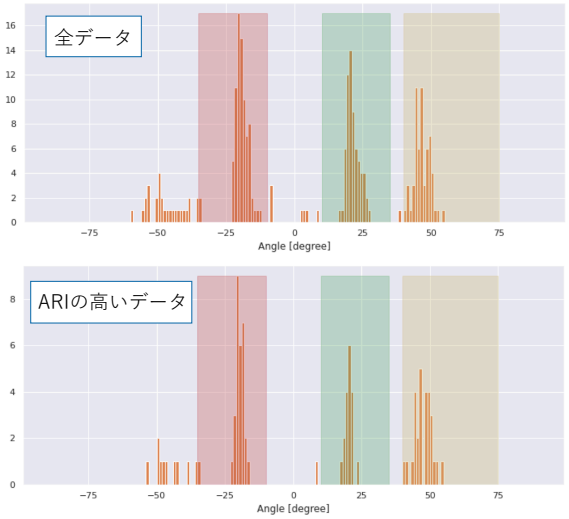


図 11 部屋方向推定性能

秒ごとに部屋を移動しながら部屋の中を自由に移動している状況で取得した音データである．

評価では，音源の位置を部屋ごとに分類する性能と，隣接する部屋の音源に関してはその方向の精度を検証した．音源の部屋分類は，クラスタリングの性能評価に一般的に用いられる調整ランド指標 (ARI: Adjusted Rand Index) を用いて評価した．部屋方向の推定精度は，マイクロフォンから見た部屋の方向との合致率を評価した．図 10 に示すようにマイクロフォンを設置した部屋はリビング，寝室，洗面所の 3 つの部屋とドアを介して隣接していることから，3.6 で k-means のクラスタ数，すなわち部屋の数は 4 とした．

図 11 に，部屋方向推定性能評価結果を示す．図は，各録音データで方向推定した結果をヒストグラムにしたものである．図で背景に色がついた角度範囲は実際の部屋方向を示しており，赤 (−35〜−10 度)，緑 (10〜35 度)，黄 (40〜75 度) はそれぞれ寝室，リビング，洗面所の方向を示している．平均 ARI は 0.725 であった．また，推定方向の正解率，すなわち実際の部屋方向である図 11 で背景色付きの角度範囲に入っていた割合は 0.850 であった．

部屋方向の推定性能を確認するため，音源の部屋の分類精度が高い場合のみを抽出した評価を行った．図 11 の「ARI の高いデータ」は，音源の部屋の分類精度が高い場合として ARI が 0.9 以上となった録音データのみを用いて方向推定を行った場合の結果を示している．この場合の推

表 2　データセットごとの部屋方向推定性能

Dataset	Mean ARI	Direction Estimation Accuracy
Bedroom DS	0.897	0.783
Kitchen DS	0.327	0.683
Washroom DS	0.746	0.967
Living DS	0.925	0.967

定方向の正解率は 0.875 であった．ARI の低い場合の結果と比べる性能が上昇していることから，音源を部屋ごとに分類する性能を高めることで部屋方向推定性能も向上できる可能性が確認できた．

次に，4 種類のデータセットごとに性能を比較した．**表 2** に，データセットごとの平均 ARI，推定方向の正解率を示す．表 2 より，キッチンデータセットを用いた場合の性能が著しく悪いことが分かる．表 1 に示したように，キッチンデータセットはマイクロフォンを設置しているキッチンにおいて 1 人の被験者が常に音を出している状況で取得したデータである．マイクロフォンを設置した部屋で取得した音はあらゆる方向からマイクロフォンに到達する可能性があるため，3.5 で示した Sound Density Map 上での音源の分離が困難となる．このために大幅に性能が低下したと考えられる．

4.3　部屋種別推定性能

部屋種別推定性能を検証するため，住宅模擬環境内で生活行動を行いながら取得した音データを用いて評価を行った．図 10a と同じ位置に 1 台のマイクロフォンを設置し，被験者 1 人がリビング・キッチン・寝室の各部屋で 30 分間自由な行動をとってもらう間に音データを取得した．そして，3.7 で述べたように無音区間を取り除くために RMS がしきい値以上となる区間をデータ分割の時間幅単位で切り出してからデータを分割した．データ分割の時間幅は 1 秒間である．

部屋種別推定に用いる学習モデルは，被験者に定められた行動を行ってもらう間に取得した音データを用いて作成した．図 10a に示した実験環境において，評価データを取得した日とは別の日にリビング・キッチン・寝室の各部屋にマイクロフォンを 1 台ずつ設置し，各部屋で起こりうる以下の行動を被験者に行ってもらい，音データを収集した．

- リビング: テレビの視聴，電話
- キッチン: 食器棚の食器の整理，皿洗い，冷蔵庫・シンク下収納の開閉
- 寝室: 寝返り，寝息を立てて睡眠

録音は全部屋の扉を開いた状態で，行動した部屋に設置されたマイクロフォンのみで行った．被験者 3 人に各行動を 120 秒間ずつ行ってもらう間に音データを収集し，120 秒間のデータをデータ分割の時間幅 1 秒間で区切って特徴量を算出して部屋種別推定学習モデルを作成した．なお，部屋種別推定学習モデルの作成では RMS に基づく無音区間の除去は行わなかった．

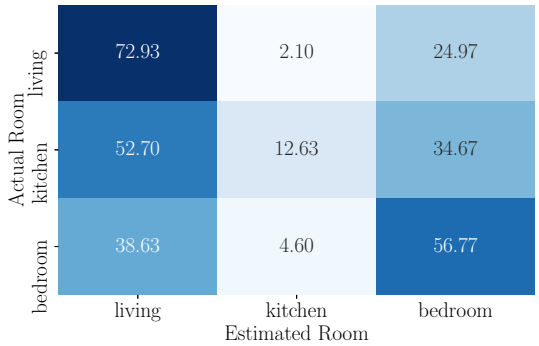


図 12　部屋種別推定結果の混同行列

本稿の評価では初期的評価として部屋方向ごとの音声合成処理は行わず，被験者 1 人が自由な行動を行っているときの音データをそのまま用いて部屋の種別を推定できるかを検証した．取得した音データに行動している部屋をラベル付けし，作成した部屋種別推定学習モデルを用いて部屋種別を推定した．

図 12 に，部屋種別推定結果の混同行列を示す．3 つの部屋の部屋種別正解率の平均値は 0.474 であった．図 12 より，リビングでは高い精度で部屋種別を推定できたことが分かる．30 分間の自由な行動では，実際には以下のような行動が行われていた．

- リビング: テレビの視聴
- キッチン: 皿洗い，食事，電子レンジの使用
- 寝室: ベッド上でのスマートフォン操作，睡眠

リビングで行った行動には学習データと共通する「テレビの視聴」という行動が行われていることから高い精度で部屋種別を推定できたと考えられる．

一方で，キッチン，寝室では十分な部屋種別推定性能を得られなかった．特に，キッチンでは学習データと共通する「皿洗い」という行動が含まれているにも関わらず推定精度は低かった．キッチン，寝室で取得した学習用データでは連続して音が出る行動が少ないことから，瞬間的な音に基づいた推定を十分に行えていないことが一因として挙げられる．このため，部屋種別推定に用いる特徴量の再検討や学習モデル作成のために収集する行動などに再検討の余地があると考えている．

5.　おわりに

本稿では，スマートスピーカを用いた家電操作において「キッチン」「リビング」などの部屋の種類 (名称) が省略された場合に操作対象機器が置かれている部屋を推定する手法として，部屋の間取りを推定した上で発話者がいる部屋を認識する手法を提案した．その実現に向けた第 1 歩として，今後のスマートスピーカには発話者がいる方向を取得

するためにマイクロフォンアレイが搭載されると想定し，マイクロフォンアレイを用いて取得した音の到来方向を解析することでどのような種類の部屋がどちらの方向に存在するかという「間取り」 推定する手法を示した．そして，1LDK 住宅模擬環境内で取得した生活行動音データを用いて初期的評価を行った結果を示した．今後は，部屋種別推定性能をさらに向上させる検討を行う予定である．

謝辞 本稿で示した研究の一部は，科研費 (JP19KT0020, JP20KK0258，JP21K11847) の助成で行われた．

参考文献

[1] Chahuara, P., Portet, F. and Vacher, M.: Making Context Aware Decision from Uncertain Information in a Smart Home: A Markov Logic Network Approach, *Ambient Intelligence*, Vol. 8309, Springer International Publishing, Cham, pp. 78–93 (2013).

[2] Chahuara, P., Portet, F. and Vacher, M.: Context-aware decision making under uncertainty for voice-based control of smart home, *Expert Systems with Applications*, Vol. 75, pp. 63–79 (2017).

[3] 永田仁史，安倍正人，城戸健一：多数センサによる音源位置の推定，日本音響学会誌，Vol. 46, No. 7, pp. 531–540 (1990).

[4] Flanagan, J. L., Johnston, J. D., Zahn, R. and Elko, G. W.: Computer-steered microphone arrays for sound transduction in large rooms, *The Journal of the Acoustical Society of America*, Vol. 78, No. 5, pp. 1508–1518 (1985).

[5] Silverman, H. F.: An algorithm for determining talker location using a linear microphone array and optimal hyperbolic fit, *Proceedings of the workshop on Speech and Natural Language*, HLT '90, USA, Association for Computational Linguistics, pp. 151–156 (1990).

[6] Wang, H. and Chu, P.: Voice source localization for automatic camera pointing system in videoconferencing, *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1, pp. 187–190 (1997).

[7] 畝田道雄，石川憲一：MUSIC による近距離音源の高分解能位置推定法に関する研究，精密工学会誌論文集，Vol. 70, No. 8, pp. 1111–1116 (2004).

[8] 畝田道雄，石川憲一：音源位置同定法としての音響ホログラフィ法と MUSIC 法の基本性能比較に関する研究，精密工学会誌論文集，Vol. 71, No. 4, pp. 517–522 (2005).

[9] Schmidt, R.: Multiple emitter location and signal parameter estimation, *IEEE Transactions on Antennas and Propagation*, Vol. 34, No. 3, pp. 276–280 (1986).

[10] 福留公利：球バッフル埋込みマイクロホンとその回折情報を利用した音源の方向及びスペクトル推定，日本音響学会誌，Vol. 44, No. 4, pp. 272–281 (1988).

[11] Tanaka, M. and Kaneda, Y.: Performance of sound source direction estimation methods under reverberant conditions, *Journal of the Acoustical Society of Japan (E)*, Vol. 14, No. 4, pp. 291–292 (1993).

[12] Warsitz, E. and Haeb-Umbach, R.: Blind Acoustic Beamforming Based on Generalized Eigenvalue Decomposition, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 15, No. 5, pp. 1529–1539 (2007).

[13] Knapp, C. and Carter, G.: The generalized correlation method for estimation of time delay, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 24, No. 4, pp. 320–327 (1976).

[14] 鈴木　敬，金田　豊：サブバンドピークホールド処理を用いた音源方向推定法，日本音響学会誌，Vol. 65, No. 10, pp. 513–522 (2009).

[15] Okamoto, T., Nishimura, R. and Iwaya, Y.: Estimation of sound source positions using a surrounding microphone array, *Acoustical Science and Technology*, Vol. 28, No. 3, pp. 181–189 (2007).

[16] Ishi, C. T., Even, J. and Hagita, N.: Using multiple microphone arrays and reflections for 3D localization of sound sources, *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3937–3942 (2013).

[17] Ribeiro, F., Zhang, C., Florêncio, D. A. and Ba, D. E.: Using Reverberation to Improve Range and Elevation Discrimination for Small Array Sound Source Localization, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 18, No. 7, pp. 1781–1792 (2010).

[18] 光来出優大，城谷知葵，石田繁巳，荒川　豊：2 つのマイクによる部屋レベル話者位置推定の検討，情報処理学会全国大会講演論文集，pp. 3:349–3:350, 4W-08 (2021).

[19] 城谷知葵，光来出優大，石田繁巳，荒川　豊：生活音からの部屋種別推定手法の検討，情報処理学会全国大会講演論文集，pp. 3:149–3:150, 1U-08 (2021).

[20] 大内一成，土井美和子：加速度と音で日々の生活行動を認識する ActivityAnalyzer，インタラクション，pp. 255–258 (2011).