

2つのマイクによる部屋レベル話者位置推定の検討

光来出優大[†] 城谷知葵[‡] 石田繁巳[†] 荒川豊[†][†]九州大学大学院システム情報科学研究所 [‡]九州大学工学部電気情報工学科

1 はじめに

近年急速に普及が進んでいるスマートスピーカは、1台で複数のIoT製品を操作することができる。複数の部屋が隣接しておりエアコンや照明など同じ製品が各部屋に設置されている環境では、あらかじめ製品が設置されている部屋を登録し、スマートスピーカに対して命令をする際にその部屋の名称を明示的に指示する必要がある。このとき、スマートスピーカ自身が命令した話者のいる部屋を推定することで「電気を消して」といったような命令でも話者の位置によって操作対象機器を推定することができる。

話者位置を推定する手法に関してはこれまでも音源定位の研究が報告されている。屋内環境では反射音の影響の軽減が課題であり、従来手法では多数のマイクを用いて複雑な信号処理を適用する必要があった。

これに対し、本稿では、2つのマイクを用いて部屋レベルで話者位置を推定する手法を提案する。スマートスピーカを用いた機器制御においては機器の設置されている部屋を推定すれば十分であることに着目し、話者のいる部屋の大きさや材質の違いから生じる音響的特徴に基づいて2つのマイクを用いて音源の特徴に基づいて話者のいる部屋の推定を行う。本稿では、複数の部屋が存在する住宅模擬施設の一室に設置したマイクから取得した音データを用いてクラスタ分析を行った初期評価について述べる。

2 関連研究

従来の屋内環境における音源位置推定手法は、ビームフォーミング法やMUSIC (Multiple Signal Classification) 法、相互相関法等様々あるが、いずれも反響や残響の影響を軽減することが課題の1つである。これらの解決策として多数のマイクを用いて複雑な信号処理により反射音の影響を軽減する手法 [1] や、3Dモデリングされた屋内環境の情報を用いて処理を行い反射音の影響を軽減する手法 [2] が提案されている。

Room-Level Speaker Position Estimation Using Two Microphones

Yudai Mitsukude[†], Tomoki Joya[‡], Shigemi Ishida[†], and Yutaka Arakawa[†][†]ISEE, Kyushu University, Japan[‡]EE/CS, Kyushu University Fukuoka, Japan[†]mitsukude.yudai@arakawa-lab.com,

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

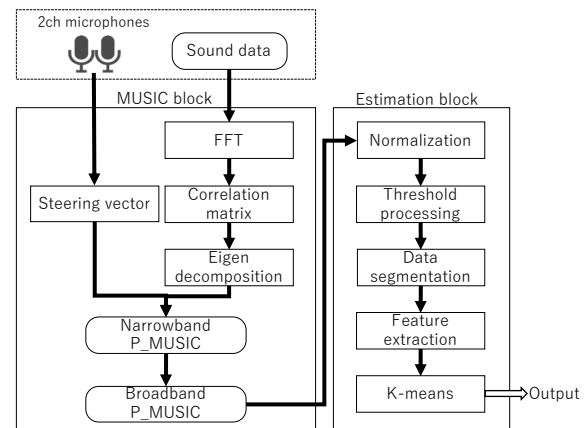
[‡]joya.tomoki@arakawa-lab.com

図1: 部屋レベル話者位置推定システムの概要

しかしこれらの手法では、必要とするマイクの数が多く信号処理や3Dモデリングなどの複雑な処理も必要となる。

3 部屋レベル話者位置推定システム

本稿では2つのマイクを用いた部屋レベル話者位置推定システムを提案する。提案するシステムは、複数の部屋が扉を介して隣接する2LDK等の住宅環境を想定している。2つのマイクはその中の一室に設置し、音データを取得する。取得した音データには、話者から直接マイクに届く直接音と壁や床に反射してマイクに届く間接音が含まれ、話者のいる部屋の大きさや材質により異なる特徴が現れることに着目し、話者のいる部屋を推定する。

図1に、提案する部屋レベル話者位置推定システムの概要を示す。マイクで取得した音データから、音源位置推定手法の1つであるMUSICアルゴリズムを用いて音の到来方向情報 (P_MUSIC) を求めた後、データ処理を行い部屋の識別を行うための特徴量を抽出する。抽出した特徴量を入力としてクラスタ分析を用いることで話者のいる部屋を推定する。提案する部屋レベル話者位置推定システムはMUSICブロックと推定ブロックから構成される。以下に、MUSICブロックと推定ブロックの処理手順について示す。

3.1 MUSIC ブロック

MUSICブロックは、音データにフーリエ変換 (FFT) を行った後にMUSICアルゴリズムを適用して

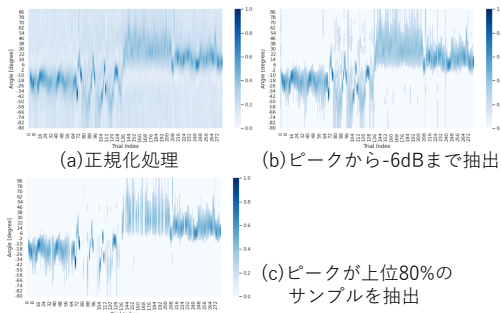


図 2: 推定ブロックの前処理手順

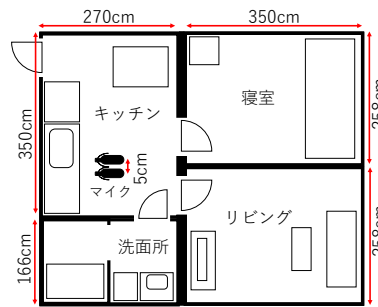


図 3: 住宅模擬施設の間取り及びマイク設置位置

	ARI 値
最大	0.977
最小	0.607
平均	0.805

図 4: 評価結果

P_MUSIC を求める。具体的には、2チャンネル音データにFFTを行いフレームごとに周波数成分を求め、チャンネル間の相関行列を求める。求めた相関行列から算出した雑音部分空間に対応する固有ベクトルとステアリングベクトルを用いて周波数ビン毎の狭帯域 P_MUSIC を求める。求めた周波数ビン毎の P_MUSIC を統合し広帯域 P_MUSIC を求める。このとき MUSIC アルゴリズムでは音源数を与える必要があるが、本稿では音源数は1であることを前提とする。

3.2 推定ブロック

図2に推定ブロックでのデータ前処理手順を示す。推定ブロックでは、MUSICブロックで求めた P_MUSIC に前処理を行った後特徴量を抽出し、それを入力としてクラスタ分析を用いて話者のいる部屋を推定する。具体的には、全サンプルの P_MUSIC に対して正規化処理を行った後に、(b) サンプル毎に P_MUSIC のピークから -6dB までを抽出する。その後、(c) サンプルのピーク値が全サンプルのピーク上位80%のサンプルのみを抽出する。

次に全サンプルを固定長のウィンドウで分割し、ウィンドウ毎に特徴量を計算する。求める特徴量は、角度の分散・角度の平均・尖度の平均・密度・1つ前のフレームの角度の平均の5種類である。これらの特徴量を入力として k-means によりクラスタ分析を行う。このとき部屋の間取りは既知とし、複数部屋の音声が入る場合を考慮してクラスタ数は部屋数+1とする。

4 評価

4.1 評価環境

図3に九州大学伊都キャンパス内の住宅模擬施設において行った評価実験の環境を示す。2台のマイクロフォン (AZDEN SGM-990) を高さ70cm、キッチンを中心に設置し、USB オーディオ (BEHRINGER UMC404HD) を用いてサンプリングレート44.1kHz、量子化ビット数16bitで音声データを取得した。評価実験では、被験者に寝室、キッチン、洗面所、リビングの順に移動してもらいながら各部屋で60秒ずつ声

を発してもらった動作を1セットとし、被験者ごとに5セット行ってもらった。被験者は24歳及び25歳の男性2名である。

評価では、各データを10秒間のウィンドウに分割しウィンドウ毎にクラスタリングを行った後にデータ全体の調整ラウンド指数 (ARI) を求めた。k-means のクラスタ数は5とし、寝室、キッチン、洗面所、リビング、不明に分類した。

4.2 評価結果

図4に評価結果を示す。評価結果より、平均ARI 0.805と高い精度で推定可能であることが確認できた。また最大ARIは0.977と非常に高い精度で部屋を推定することが可能である一方で最小ARIは0.607と精度が低くなるのが分かる。これはマイクを設置しているキッチンにおいて音源の方向が変化しやすい点、キッチンが他のどの部屋とも隣接しているため他の部屋の入り口付近に音源がある場合誤った推定結果になりやすい点が原因として考えられる。

5 おわりに

本稿では、2つのマイクロフォンを用いて部屋レベルで話者位置を推定する手法を提案した。評価の結果より、平均ARI 0.805という高い精度で話者のいる部屋を推定することができた。今後の課題として、反響が大きく角度の推移が大きいマイク設置部屋の影響を軽減する手法について検討を行う。また、音源数が複数になった場合での話者のいる部屋の推定についても検討を行う予定である。

謝辞

本研究の一部はJSPS 科研費JP19KT0020の助成で行われた。

参考文献

- [1] 湯浅 他. MUSIC 音源方向推定法におよぼす反射音の影響とSBPHP処理の検討, 日本音響学会春季講演 (2018).
- [2] Ribeiro, F. et al.: Using reverberation to improve range and elevation discrimination for small array sound source localization, *IEEE Trans. Audio, Speech, Language Process.*, Vol. 18, No. 7 (2010).