神戸大学経済経営研究所
Research Institute for Economics and Business Administration, Kobe University

■コラム　RIEB ニュースレターNo.195　2019 年 2 月号

音源推定研究の簡単紹介

神戸大学　経済経営研究所
助教　陳　金輝

AIの核心をなす機械学習研究は画像，音声処理へ多く活用されている．画像処理の研究はマスコミに頻繁に報道される．しかし，音声処理研究に関してニュースで紹介されることはかなり少ない．音声は，人間にとって最も身近な情報のひとつであり，人間同士のコミュニケーションをはじめ，メディアによる情報の伝達など，多くの場面で利用されている情報である．ここで，我々は音声から「何と言っているのか」を表す言語情報だけではなく，「誰が」，「どのような気持ちで」，「どのような環境下で」話をしているのか，すなわち話者・感情・環境といった多くの情報を読み取ることができる．そしてこれらの情報は，特に人間同士のコミュニケーションを円滑に進める上で重要な役割を担っている．

音声から読み取れる多くの情報の中には，「どこから話されているのか」を表す音源位置・方向の情報も含まれる．機械学習システムはこの位置や方向の情報をもとに，話者の区別，特定話者動作の表現（e.g.，仕草特徴記述子のモデリング）や，話者の位置の探索を行うことができる．また，目的の方向からの音声だけを意識して聞こうとすることで，雑音の多い環境下でも目的の音声を認識することができる．このため、音源情報は音声処理の領域において重要である。現在行われている音声の研究の中には，この音源位置・方向の情報を用いて音声の情報の豊富化や，音声インターフェースのロバストネス（robustness）の向上を行う研究が多く存在する．

音声情報の豊富化に関しては，音源位置の情報を用いることで発話者を特定し，議事録の作成など，会話状況の詳細な分析を行う手法が提案されている．また，音源方向の情報から話者が交替している時間フレームを検出し，その情報をもとに，システムに入力された音声がシステムへの要求なのか，単なる雑談なのかを判別する手法も提案されている．ロバストネスの向上に関しては，遅延和アレーや適応型アレーのように，目的音声の方向に指向性を形成する，あるいは雑音方向に死角を形成することで目的音声を強調や雑音の抑圧を行う手法が多く提案されている．

これまでに多くの音源位置推定の研究がされてきたが，これらの研究では，マイクロホンアレーと呼ばれる複数のマイクロホンを用いて，音源位置を推定する手法がほとんどである．マイクロホンアレーに基づく手法では，音声が各マイクロホンに到達する時間が音源の位置によって異なる点に着目し，各マイクロホンにおける観測信号間の時間差（位相差）を用いて音源位置を推定している．一方，単一マイクで処理を行おうとする試みは雑音抑圧や音源分離などの分野でも研究されてお

り，関連手法も複数提案されている．しかしながら，単一マイクロホンで音源の位置や方向を推定しようとした場合，従来手法のような信号間の位相差といったマイク間の情報が使えないため，別の情報を用いた音源位置・方向推定手法が必要である．

単一マイクロホンで音源の位置や方向を推定する研究はかなり困難なタスクであり，学術価値が高いものとされている．しかしながら，マイク一つのみで音源位置を推定する方法論は未だ確立されておらず，単一マイクロホンによる音源位置推定は非常に困難な試みとされていた．試みた有効な方法として，音声の持つ音響伝達特性が音源位置や方向に依存する点に着目し，音響伝達特性を用いることで音源の位置や方向を単一マイクロホンのみで推定する手法について提案し，そのアプローチとして大きく二つの枠組みをなす．一つは，マイクを中心に回転するパラボラ反射板を用いて，反射板の回転により変動した音響伝達特性を検出することで，音源の方向を推定する，アクティブマイクロホンによる音源方向推定法である．もう一つは，部屋の音響伝達特性が音源位置に依存して変化する点に着目し，音響伝達特性を識別することで，音源の位置を推定する手法である．また，後者の枠組みを応用した手法として，シングルチャネルによる複数話者の音源位置推定手法，話者の頭部方向の推定システムにも活用することができる．

マイクによって観測される信号は，原音声（クリーン音声）と音響伝達特性（インパルス応答）が畳み込まれた残響信号である．そのため上述システムでは，残響信号からクリーン音声成分を取り除き，音響伝達特性成分のみをブラインドで推定する必要がある．この手法では，あらかじめ話者のクリーン音声をGaussian Mixture Model (GMM)あるいはHidden Markov Model (HMM) でモデル化しておき，これをクリーン音声成分の事前知識として活用することで，観測信号から音響伝達特性を確率的に推定している．そして，推定された音響伝達特性を用いて音源の位置や方向を推定する．しかしながら，観測された信号から音響伝達特性を推定する際の誤差が，その後の音源位置・推定の性能の低下に繋がっており，従来のマイクロホンアレーに並ぶ精度を得るためには，より正確な音響伝達特性の推定が必要であると考えられる．また，未学習の話者や位置，部屋環境の変化といった今後解決しなければならない問題が多く存在している．これらの問題を解決するために，不特定話者モデルを用いて話者の適応を行いながら音響伝達特性を推定する手法や，未学習の位置や環境におけるオンライン学習・適応といった手法についても今後検討する必要がある．

※関連内容論文をトップジャーナルIEEE Transactions on Cyberneticsに投稿する予定．