博士（情報科学）　須山（荒木）章子

学位論文題名

Convolutive Blind Speech Separation with Independent Component Analysis and Sparse Component Analysis

（独立成分分析およびスパース成分分析によるブラインド音声信号分離）

学位論文内容の要旨

我々が生活している環境は常に音に満ちているが、このような中で、例えばロボット（コンピュータ）に音声認識をさせようとしても、目的話者音声以外の雑音のために音声認識機は性能を十分に発揮することはできない。また、我々人間にとっても、例えば遠隔会話（携帯電話やテレビ会議システムなど）の相手がざわざわした場所にいると、会話がしづらいといったこともある。混合音の中から目的音声を分離抽出することは、コンピュータと人との対話・人と人との対話双方で重要な課題である。

混合音からの目的音声抽出を実現する１つの方法として、本論文では、ブラインド音源分離に着目する。ブラインド音源分離とは、音源や混合の情報（話者・話の内容・話者位置・部屋の残響条件など）を知ることなく、観測された混合信号のみから音源信号を推定し復元する技術である。

ブラインド音源分離の有力な手段としては独立成分分析がある。これは複数の混合信号から、それを形成する複数の互いに独立な信号を推定する、統計的教師なし学習手法であり、脳波信号などで既に分離に成功している。そのため独立成分分析は万能の感があったが、音声信号（広帯域・非定常）が、残響のある実際の環境で混ざってしまった場合のような複雑な問題に対する検討はなされていなかった。

そこで本論文ではまず、音声信号に独立成分分析を適用するために、統計学的な理論・歴史を持つ独立成分分析に対して音響信号処理的な観点からの解釈を与えた。筆者は、独立成分分析が、マイクロホンアレイシステムの代表例である適応型ビームフォーマと同じメカニズムで動作すること、すなわち、不要信号に死角を形成し、目的信号のゲインを保って出力するマイクロホンアレイに等しいことを示した。これにより、独立成分分析により音源分離を行う際にも適応型ビームフォーマと同じように、帯域に適したマイク間隔を用いたり、周波数領域独立成分分析固有の問題である出力順序や出力振幅の不定性の解決に、マイクロホンアレイ系の指向特性を用いたりすることができるようになった。

筆者はまた、残響のある環境下で混合された音声に対して周波数領域での独立成分分析を適用し、周波数分析フレーム長と性能との関係を明らかにした。ここでは、分析フレーム長が短いと長い残響に対応できず、フレーム長が長いと結果的に各周波数での時系列サンプル数が少なくなるために独立成分分析に必要な統計量の計算に支障をきたす、というトレードオフが明らかとなった。

このトレードオフを緩和する１つの方法として、サブバンド領域における独立成分分析を用いたブラインド音源分離を提案した。これは信号を適切な帯域分割数へサブバンド分解し、各帯域で時間領域の独立成分分析を用いて音源分離を行うものである。これにより、各帯域では

統計量の計算に十分なサンプル数を保持できるようになり、また時間領域の独立成分分析にて各帯域である程度の長さを持つ分離フィルタを使えることから残響への対処もできるようになった。そのため、サブバンド領域における独立成分分析は周波数領域での独立成分分析より高い性能を達成することを確認できた。

独立成分分析によるブラインド音源分離は、音源数がマイク数と同じか少ないことを仮定したものであり、音源数がマイク数以上になると性能は著しく劣化する。そこで、音源数がマイク数以上でかつ、マイクが自由に配置されている場合にも適用可能な、ブラインド音源分離手法を提案した。

これは、音声信号のスパース性を積極的に利用した、スパース成分分析に基づく手法である。音源がスパース信号であることを仮定することで、各時間周波数には音源が高々1 つしか存在しないという仮定を置くことができ、これより、各時間周波数でアクティブな信号の位置情報を推定・クラスタリングする。最終的には、各クラスタのメンバを時間周波数領域バイナリマスクにて抽出することにより分離が達成される。この方法も、独立成分分析による分離と同様に、音源や混合の情報を必要としないブラインドな手法となっている。

提案法は、無響環境を仮定した方法であるが、普通の会議室程度の残響のある部屋でも動作することが確認された。

スパース成分分析によるブラインド音源分離では、音源数がマイク数以上でも音源分離が可能であるが、時間周波数領域バイナリマスクにより分離信号に不連続性が発生し、耳障りな非線形ノイズを発生させる。これを低減させるために、独立成分分析と、スパース成分分析とを併用するブラインド音源分離手法を提案した。この主眼は、時間周波数マスクの利用を最小限にし、主な分離は線形変換である独立成分分析で行うことである。具体的には、まず、時間周波数マスクにより観測混合信号中の音源数をマイク数以下まで減らし、そのうえで独立成分分析を適用する。

これは簡便な方法であるが、音源数がマイク数以上の場合にも高い分離性能が得られた。

以上の研究により、残響のある実際の環境で混合された音声信号のブラインド音源分離の基礎技術が確立された。

学位論文審査の要旨

主　査　　教　授　　牧　野　昭　二

副　査　　教　授　　荒　木　健　治

副　査　　教　授　　長谷川　美　紀

副　査　　教　授　　山　本　　強

副　査　　教　授　　宮　永　喜　一

学位論文題名

Convolutive Blind Speech Separation with Independent Component Analysis and Sparse Component Analysis

（独立成分分析およびスパース成分分析によるブラインド音声信号分離）

混合音からの目的音声抽出を実現する１つの方法として、本論文では、ブラインド音源分離に着目している。ブラインド音源分離とは、音源や混合の情報（話者・話の内容・話者位置・部屋の残響条件など）を知ることなく、観測された混合信号のみから音源信号を推定し復元する技術である。

本論文ではまず、音声信号に独立成分分析を適用するために、統計学的な理論・歴史を持つ独立成分分析に対して音響信号処理的な観点からの解釈を与えた。筆者は、独立成分分析が、マイクロホンアレイシステムの代表例である適応型ビームフォーマと同じメカニズムで動作すること、すなわち、不要信号に死角を形成し、目的信号のゲインを保って出力するマイクロホンアレイに等しいことを示した。これにより、独立成分分析により音源分離を行う際にも適応型ビームフォーマと同じように、帯域に適したマイク間隔を用いたり、周波数領域独立成分分析固有の問題である出力順序や出力振幅の不定性の解決に、マイクロホンアレイ系の指向特性を用いたりすることができるようになった。

筆者はまた、残響のある環境下で混合された音声に対して周波数領域での独立成分分析を適用し、周波数分析フレーム長と性能との関係を明らかにした。ここでは、分析フレーム長が短いと長い残響に対応できず、フレーム長が長いと結果的に各周波数での時系列サンプル数が少なくなるために独立成分分析に必要な統計量の計算に支障をきたす、というトレードオフが明らかとなった。

このトレードオフを緩和する１つの方法として、サブバンド領域における独立成分分析を用いたブラインド音源分離を提案している。これは信号を適切な帯域分割数へサブバンド分解し、各帯域で時間領域の独立成分分析を用いて音源分離を行うものである。これにより、各帯域では統計量の計算に十分なサンプル数を保持できるようになり、また時間領域の独立成分分析にて各帯域である程度の長さを持つ分離フィルタを使えることから残響への対処もできるようになった。

独立成分分析によるブラインド音源分離は、音源数がマイク数と同じか少ないことを仮定したものであり、音源数がマイク数以上になると性能は著しく劣化する。そこで、音源数がマイク数以上でかつ、マイクが自由に配置されている場合にも適用可能な、スパース成分分析によるブラインド音源分離手法を提案している。

スパース成分分析によるブラインド音源分離では、音源数がマイク数以上でも音源分離が可能であるが、時間周波数領域バイナリマスクにより分離信号に不連続性が発生し、耳障りな非線形ノイズを発生させる。これを低減させるために、独立成分分析とスパース成分分析とを併用するブラインド音源分離手法を提案し、音源数がマイク数以上の場合にも高い分離性能を得ることを示している。

これを要するに、著者は音声情報処理分野で混合音から、特定音源からの目的音声を抽出する手法、特に音源に関する情報を与えないで分離処理を行うブラインド音源分離の手法について新たな知見を得たものである。よって著者は北海道大学博士（情報科学）の学位を授与される資格あるものと認める。