様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　２７　年　　５　月　１９　日現在

機関番号：　１４４０１
研究種目：　基盤研究(C)
研究期間：　2012～2014
課題番号：　２４５００２０４

研究課題名（和文）ＭＡＰ推定ノイズ除去と単一話者区間推定に基づくブラインド音源分離に関する研究

研究課題名（英文）MAP estimation-based noise suppression and blind source separation using single voice activity detection

研究代表者

川村　新（KAWAMURA, Arata）

大阪大学・基礎工学研究科・准教授

研究者番号：６０３６２６４６

交付決定額（研究期間全体）：（直接経費）　　3,500,000円

研究成果の概要（和文）：本研究では，2つのマイクロホンを用いて，複数の音源を分離する方法について検討した．実環境においては，環境ノイズと残響が分離性能を劣化させる要因となるため，音源分離の他に，これらを取り除く技術が必須となる．そこで提案法では，単一話者区間（いずれか一人のだけが発話している区間）を検出し，この区間を利用して各発話者の位置を特定し，音源分離を実行する．また，MAP（maximum a posteriori）推定ノイズ除去が音源分離の後処理として有用なことを示す．本手法が残響およびノイズに対して頑健であることを実環境における音源分離実験により示した．

研究成果の概要（英文）：A blind source separation method using two microphones has been investigated. In a practical situation, environmental noise and reverberation exist. Since they degrade source separation quality, we have to remove such undesired effects. To establish an effective blind source separation method, we employ a single voice activity detector which detects single talk segments. These segments give the target source locations. Addition to it, a MAP (maximum a posteriori) estimation-based noise suppressor is introduced as a post-processor for improving the speech quality of the separated signals. Test speech signals are transmitted from loudspeakers and captured at a stereo microphone in a practical reverberant environment. Simulation results showed that the observed speech signals are effectively separated by the proposed method.

研究分野：　音声信号処理

キーワード：　単一話者区間　音源分離　ノイズ除去　事後確率最大化　統計処理

１．研究開始当初の背景

ロボットやカーナビゲーションシステムの音声認識装置や，携帯電話による通話では，環境ノイズや多数の音声の中から，必要となる音声だけを取り出すことが望ましい．さらに，観測信号を個々の音源に分離できれば，複数音声の同時認識や，テレビや映画などで必要な音声の取捨選択が可能となり，広範な応用が期待できる．小型機器への搭載，製造コスト，さらには人間や動物の耳の数を考慮すれば，使用マイクロホンの数を 2 以下とすることが望ましい．代表的な 2 マイクロホン音源分離法として，DUET (Degenerate Unmixing Estimation Technique)[1]が知られている．DUET では，各音源が同じ時刻に同じ周波数成分を共有しないという仮定 (W-DO 仮定) の下で，音源分離を実行する．しかし，環境ノイズが存在する場合，W-DO 仮定が成立せず，音源分離性能は劣化する．そこで，筆者らが研究を続けてきた MAP(最大事後確率) 推定によるノイズ除去法を，DUET の前処理として導入し，ノイズの影響を減らす．さらに本研究では，ある時刻にひとつの音源だけが観測される，単一話者区間が存在すると仮定する．このとき，音源の到来方向を表すベクトルに直交するベクトルが作成できる．このベクトルに観測信号を射影(直交軸射影) すれば，対応する音源を 1 つ消去できる．直交軸は音源数だけ得られるから，2 音源であれば射影後の信号がそのまま分離信号となる．一方，音源が 3 つ以上のときは，W-DO 仮定を利用して，直交軸射影後の任意の 2 つの信号の振幅比が 0:1 となる周波数を抽出すれば，音源分離が実現できる．筆者は，ノイズ除去と直交軸射影を個別に研究し，それぞれの性能を評価してきた．これまでの研究成果として，筆者が提案した適応 MAP 推定によるノイズ除去法は，従来の Lotter らの MAP 推定[2] よりも，ノイズ除去性能が高くなることを確認しており，論文発表を行っている．また，単一話者区間を利用した直交軸射影の有効性についても，すでに学会で発表している．ただし，これらの結果は，音源の混合過程を，瞬時混合と呼ばれる単純な過程として得たものである．よって，より現実の問題に近い，畳み込み混合モデルへ直交軸射影法を拡張することが解決すべき課題である．さらに，MAP 推定に基づくノイズ除去と，単一話者区間を利用した直交軸射影を融合した，提案システム全体の演算量を評価し，オンライン処理実現のために改善すべき課題を明確にする．そして，本課題の最終目標として，提案法の実環境における性能評価を行う．

２．研究の目的

本研究の目的は，複数の音声とノイズが混在する信号からノイズを除去し，さらにそれぞれの音声を分離することである．筆者は，使用するマイクロホンを人間の耳と同数の 2 つだけに限定して，この問題の解決を目指す．まず，筆者らがこれまで個別に研究してきた，ノイズ除去と直交軸射影を融合し，瞬時混合モデルにおける全体のシステムを構築する (図 1). ここで, 周囲ノイズと音声の SNR は，携帯電話で対象とされる 6dB 以上とし，音源の数は 3～5 程度の環境を想定する．標準的なノイズ除去では，ノイズパワーを 1/4 にできれば十分に有効性が認められる．これは 6dB の SNR 改善に相当する. 瞬時混合モデルにおいては，問題設定が単純なのでノイズ除去で 10dB 以上, 音源分離性能で 10dB 以上の改善を目指す．このとき，観測信号として得られた信号に対して，全体で 20dB 以上の SNR 改善となる．次に，瞬時混合モデルで構築された提案システムを，現実の環境に近い畳み込み混合モデルに拡張する．そして，ノイズ除去において 6dB 以上，音源分離性能で同じく 6dB 以上の SNR 改善を目標とする．また，単一話者区間を利用する直交軸射影により，DUET に従来必須であったヒストグラム作成，ピークサーチが不要になり，演算量を約 70 %以下に削減することが可能である．よって，演算量削減も本研究の重要な目的のひとつである．

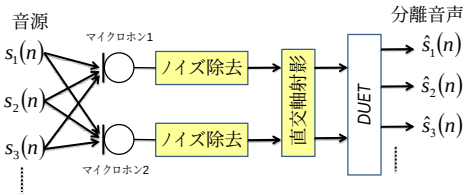

図 1　ノイズ除去と音源分離の融合

３．研究の方法

以下の方法に従って，各研究項目を遂行した．

（１）単一話者区間に基づく直交軸射影法の確立

直交軸は，音源がひとつで，残りの音源が休止している単一話者区間において得られる．この区間内では，2 つのマイクロホンで得られる観測信号は，定数倍を除いて全く同じ波形となる．このとき，両者の比は混合行列のいずれかの列ベクトル，すなわち音源の到来方向を表わすベクトルを与える．これと直交するベクトルが張る軸が直交軸となるから，直交軸は瞬時に得られる．最も重要な問題は，音源がひとつとなる区間の検出である．この区間は，逆に，2 つの観測信号の比が定数となる区間から推定することができる．しかしながら，実際には音源がひとつであっても，マイクロホンのノイズ等で観測信号の比にゆらぎが生じる．そこで，適切なしきい値を求め，比のゆらぎを吸収する必要がある．ゆらぎを適切に吸収できれば，原理通り直交軸射影による 1 音源の除去が実現できる．当然ながらゆらぎの吸収は，音源分離の分解能を劣化させる．そこで，これらのト

レードオフ関係を明らかにしておき，実際の運用時の設計指針とする．

（２）直交軸射影と ICA の関係の明確化

2 音源の場合を考えると，上述の直交軸者泳法と ICA が探索する基底軸は同じである．ICA により得られた基底軸においても，それぞれの軸に信号を射影すると，分離された各音源が得られる．ICA は独立性を評価して軸を探索することに対し，直交軸射影では音源がひとつとなる区間から軸を探索する．また，音源数が増えた場合もマイクロホン数を増やせば，両者は同じ結果を与える．このように両者は深く関係しており，これらの関連性を理論的に示すことは，ICA を基礎とする手法に対しても有益である．また，上記の課題遂行と並列して，関連会議への参加により，最新の 2 マイクロホン音源分離法について他の研究者と積極的な意見交換を行う．

（３）システムの構築と評価

ノイズが定常的に存在する環境では，DUET で必要となる W-DO 仮定が成立しないため，音源分離性能が著しく劣化する．これを解決するため，MAP 推定ノイズ除去[3]を導入する．MAP 推定ノイズ除去により W-DO が満たされ，DUET が効果的に実行されると期待できる．しかしながら，ノイズ除去処理の影響で，所望音声のある程度の劣化は避けられない．MAP 推定ノイズ除去は，ノイズ除去量や音質に関してパラメータによる調整が可能であるが，音源分離結果に与える影響について慎重に評価する必要がある．また，さらに音源分離性能を改善できる付加的な手法についてサーベイを行い，積極的に導入を検討する．ただし，この段階では，実用化レベルの高性能な分離結果を得ることに専念し，音源分離ための演算量については言及しない予定である．

（４）実環境への対応とシステムの洗練化

最後に，演算量削減を中心としてシステムの洗練化を行う．今回導入予定の MAP 推定ノイズ除去法は，現時刻の観測信号スペクトルと，事前推定したノイズスペクトルを利用して，音声スペクトルを推定するものである．しかし，提案システムにおける 2 つの観測信号に含まれる音声スペクトルは類似性をもつので一方の結果を利用できる可能性が高い．また，方向性をもつノイズではなく，拡散性ノイズが音声に重畳する場合，2 つのマイクロホンで観測されるノイズは同じ特性をもつ．したがって，2 系列のノイズ除去システムを独立に動作させるのではなく，相互に関連させることで，1 系列に近い状態まで演算量を削減できる可能性がある．ところが，いくつかの実環境音の分析結果から，主に残響がノイズとしてふるまい，分離性能が劣化することが確認できた．そこで，残響に強い従来方式[4]の技術を提案法に取り入れることにより，実環境における音源分離性能を改善するように研究の方向性をやや修正する．さらに，従来法[4]は演算量が膨大であるため，その演算量削減について検討する．

＜引用文献＞

[1] O. Yilmaz and S． Rickard,「Blind separation of speech mixtures via time-frequency masking」, IEEE Trans. on SP, vol.52, no.7, pp.1830{1847, July (2004).

[2] T. Lotter and P. Vary, 「Speech enhancement by MAP spectral amplitude estimation using a Super-Gaussian speech model」, EURASIP Jurnal, vol. 2005, no. 7, pp.1110{1126, July (2005).

[3] Y. Tsukamoto, A. Kawamura and Y. Iiguni, 「Speech Enhancement Based on MAP Estimation Using Variable Speech Distribution」, IEICE Trans. Fundamentals, Vol.E90-A, No.8, pp.1587－1593 (2007).

[4] H. Sawada, S. Araki, and S. Makino, 「 Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment」, IEEE Trans.ASLP, vol.19, no.3, pp. 516－527 (2010).

４．研究成果

（１）最初に，「単一話者区間に基づく直交軸射影法の確立」と，「直交軸射影法と ICA との関係を明確にすること」の 2 つに取り組んだ．まず，直交軸射影法を，瞬時混合と時間遅延を含む混合のそれぞれに対して導出した．直交軸は，音源がひとつで，残りの音源が休止している単一話者区間において得られる．この区間内では，2 つのマイクロホンで得られる観測信号は，定数倍を除いて全く同じ波形となる．よって，両者の比は混合行列のいずれかの列ベクトル，すなわち音源の到来方向を表わすベクトルを与える．これと直交するベクトルが張る軸が直交軸である．ここで，重要な問題は，音源がひとつだけとなる区間の検出である．この区間は，2 つの観測信号の比が定数となる区間として推定した．しかしながら，実際にはマイクロホンノイズ等の影響で観測信号の比にゆらぎが生じる．そこで，あるしきい値により，比のゆらぎを吸収する方法を検討した．結果から，ゆらぎの吸収性能と，音源分離性能にトレードオフの関係があることがわかった．今回は実験的にしきい値を決定したが，将来的には適応設定する必要があろう．次に，上述の「直交軸」と，ICA で得られる「基底軸」の関係を明らかにした．両者は，いずれも同一の単位ベクトルを定数倍したものとして表現できる．ICA では，基底軸に信号を射影することで，音源分離を実現する．我々の直交軸射影も全く同じ原理で音源分離を実行する．異なる点は，ICA が独立性を評価して軸を探索することに対し，直交軸射影では音源がひとつとなる区間から軸を探索する点である．我々の方法は必ずしも独立性を必要としないという点において，ICA よりも汎用性が高いといえる．その他，関連会議への参

加により，最新の 2 マイクロホン音源分離法について他の研究者と積極的な意見交換を行い，以降の研究進展における重要なヒントを得た．

（2）次に，MAP 推定ノイズ除去法と，直交軸射影を DUET に組み込み，システム全体を構築した．ノイズが定常的に存在する環境では，DUET で必要となる W-DO 過程が成立しないため，音源分離性能が著しく劣化する．そこで筆者はこの問題を，図 1 に示したように，MAP 推定ノイズ除去を音源分離の前段に導入し，解決を試みた．これは，MAP 推定ノイズ除去により，W-DO 仮定が満たされ，DUET が効果的に実行されることを期待するものである．データベースの音声にノイズを付加した信号に対し，構築したシステムでノイズ除去および音源分離を実行した．評価結果から，予想に反して分離性能が改善しないことがわかった．原因を調査したところ，ノイズ除去後の信号から音源方向を特定できておらず，音源分離に失敗していた．ノイズ除去後のスペクトルから生成したヒストグラムを図 2 に示す．波形では MAP 推定によるノイズ除去が達成されているにもかかわらず，ヒストグラムからは音源方向に関する情報が読み取れない．これは，位相スペクトルの劣化を改善できないことが原因であると考えられる．そこで，観測信号にノイズ除去を行わず，音声が優勢となるスペクトルだけを用いて音源方向を特定した．音声が優勢となるスペクトルから生成したヒストグラムを図 3 に示す．これより，3 つの音源に対するピークが確認できる．これらのピークを利用し，ノイズが含まれたままの観測信号に対し，音源分離を実行する．分離した音声にはノイズが含まれているので，これに対して，MAP 推定によるノイズ除去法を適用する．この構成を図 4 に示す．結果として，音源分離を実行した後にノイズ除去を行った方が，高い分離性能が得られることがわかった．

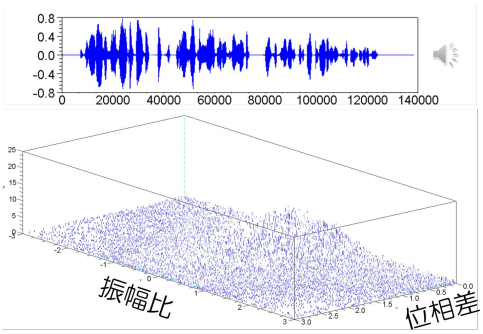

図 2 ノイズ除去後のスペクトルから生成したヒストグラム．ピークが確認できない．

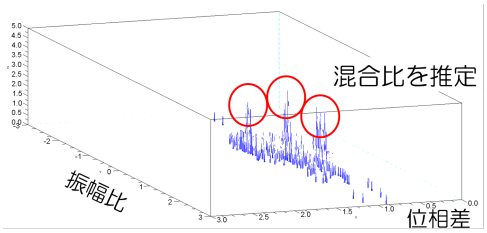

図 3 ノイズ除去を行わず音声が優勢なスペクトルから生成したヒストグラム．3 つのピーク（混合比）が確認できる．

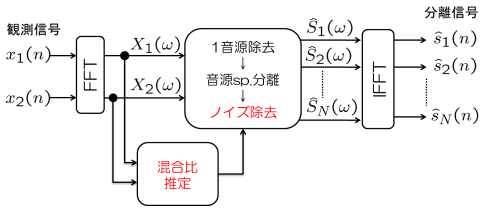

図 4 ノイズ除去を音源分離の後にした構成

（3）2 年目までの成果により，単一話者区間に基づく音源分離を実施した後に，MAP 推定ノイズ除去を適用することが望ましいことがわかった．最終年度では，実環境での提案法の利用を重視し，残響および環境ノイズに対応できる，音源分離技術の開発を主体とした． 2 年目までの方法では，部屋の残響が十分小さいことを想定したシミュレーションにおいて，単一話者区間に基づく音源分離法が有効であることを確認していた．しかし，提案法は，原理的に強い残響特性が存在する環境へは対応できない．実際，研究室において収録した音声に対しては，いわゆる環境ノイズは小さいものの，残響がノイズとしてふるまい，音源分離性能が著しく劣化した．そこで，残響に頑健な従来法[4]の技術を提案法に導入することにした．このため，リアルタイム処理が可能であった提案法を，いったんバッチ処理方式に作り変えた．これは，単一話者区間において，残響の周波数特性をベクトル化し，すべての単一話者区間で平滑化することにより，残響およびノイズ耐性に優れた音源分離を実現する方法である．本研究における提案法の最終構成を図 5 に示す．

原理的には従来法[4]は，提案法と同様の結果が得られるものの，ベクトルで音源分離を実行する提案法に比べて演算量が膨大となる．実環境において収録した 2～6 名の混合音声（各 200 通り）に対して，音源分離実験を行った．収録環境を図 6 に示す．ここで，スピーカ 6 台とステレオマイクロホン 1 台を使用した．音源分離結果を図 7 に示す．ただし，評価は SIR（Signal to Interference Ratio）で実施し，大きいほど性能が良いことを示している．結果から，提案法は従来法

とほぼ同等の音源分離性能を示した．特に，4 人までの少ない音源数に対しては，従来法を上回る結果が得られた．また，演算時間に関しては，提案法が，従来法の 5.5%～35%となり，明らかな有効性が示された．

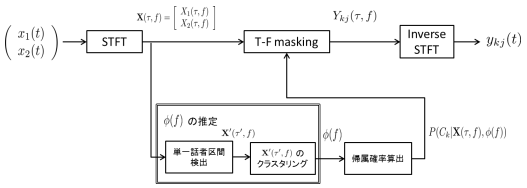

図 5 残響にロバストな音源分離法（提案法）

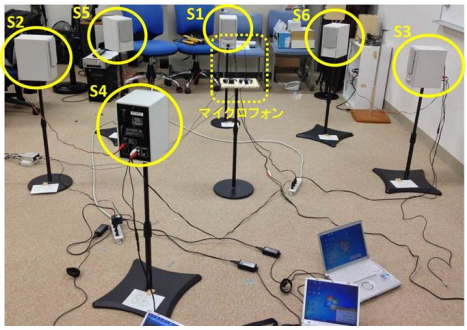

図 6 実験環境

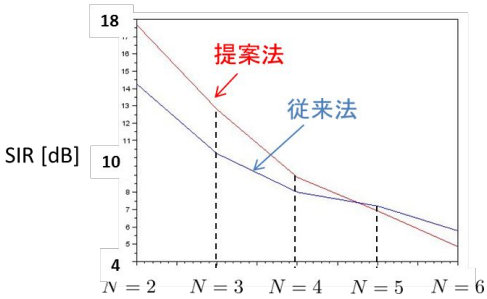

図 7 評価結果．SIR が大きいほど高性能．N は話者数を表す（200 通りの混合音声に対する平均値）．

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔学会発表〕（計　1 件）

松田篤史，川村新，飯國洋二，単一話者区間に基づいた残響下での 2 チャンネル音源分離，電子情報通信学会技術報告，SIP2014-122，2015，pp.55-60.

６．研究組織
(1)研究代表者

川村　新　（KAWAMURA，Arata）
大阪大学・大学院基礎工学研究科・准教授
研究者番号：６０３６２６４６