■2 群（画像・音・言語）－ 6 編（音響信号処理）

2 章 音源分離

（執筆者：浅野　太）［2011 年 11 月受領］

■概要■

われわれの生活している環境には様々な音が存在する．その中から所望の音を抽出する技術が求められている．例えば，音声認識を行う際には，周囲の不要な音を取り除いて，認識対象となる音声のみを収音したい．また，存在する複数の音それぞれを分離測定する技術も望まれている．会議音声記録などがその一例となる．ここでは，そのような技術を音源分離技術と呼び，その代表的な手法を紹介する．

一つのマイクロホンで収音された 1 チャネル信号から，不要な信号を取り除く技術は雑音抑圧技術と呼ばれ，4 章で解説を行う．本章では，マイクロホンアレー（複数のマイクロホンを用いた収音装置）を用いて収音した複数チャネルの信号を利用して音源分離を行う技術を紹介する．複数マイクロホンを用いれば，音の空間的情報（音の到来方向や音源距離など）を得ることができる．そして，複数の音の空間的な性質の違いに基づいた分離を行うので，類似性の高い信号（例えば二つの音声）であっても分離が可能である．

空間的な分離は，適切な指向特性を形成することによって実現されるが，その前提条件によって技術が分類される．第 1 は，目的とする方向を利用者が定めて収音する場合である．例えば，目的音源の方向が既知の場合や，目的とする方向の（例えば，カメラの向いている方向の）音を収音したい，というような場合である．この場合には，アンテナやソナーなどの分野でも利用されているビームフォーミングという技術が利用されている．ビームフォーミングでは目的音方向の感度を確保しつつ，目的音方向以外の感度を低下させる．

一方，目的音の方向が未知の場合には，独立成分分析が利用できる．この技術は，音源信号の独立性を利用して複数信号の分離を行うものであるが，実際に分離を行うための操作は指向性制御と等価なものである．この技術は目的音方向などの情報が未知でも動作可能なことから，ブラインド信号分離と呼ばれている．

また，音声信号は時間－周波数領域において存在がまばらなスパース性と呼ばれる性質をもっている．すなわち，複数の音声が同時に存在しても，ある時刻のある周波数においては一つの音声のみが存在する場合が多い．この性質を利用することで，より高機能な指向性制御が可能となる．

【本章の構成】

本章では，まず 2-1 節で，ビームフォーミングの原理と基本問題，及び代表的な手法として固定ビームフォーミングならびに適応ビームフォーミングについて説明する．次に 2-2 節で，ブラインド信号分離について説明し，2-3 節では，信号のスパース性を利用した音源分離技術の説明を行う．

■2群 － 6編 － 2章

2-1 ビームフォーミング

（執筆者：宝珠山治） ［2011年11月受領］

マイクロホンアレーなどを用いて指向性（方向に関する選択性）を制御する信号処理技術をビームフォーミング，それを行うシステムをビームフォーマと呼ぶ．音源からマイクロホンへの音波伝搬がそれぞれ異なることに基づき，遅延及びフィルタにより位相や振幅を制御した信号同士を干渉させ，特定の方向からの信号を強調あるいは低減する．音を対象としたビームフォーミングでは，扱う信号の周波数，波長の範囲が広い点が，比較的狭帯域の信号を扱うアンテナアレーと異なる．ビームフォーミングは大きく分けて，環境によらず固定した信号処理を行う固定ビームフォーミングと，環境に適応して処理を可変する適応ビームフォーミングがある．

本節では，まず，ビームフォーミングの基本原理及び基本問題について述べる．続いて，固定ビームフォーミング技術を紹介し，最後に，適応ビームフォーミングについて述べる．

2-1-1 ビームフォーミングの原理

まず**図 2･1** を用いて，最も簡単な2マイクロホンの場合を例に，基本原理を説明する．特性が全く等しい2個の全指向性マイクロホンを間隔 d で配置し，これらに対して，平面波が方向 θ から到来する状況を考える．この平面波は各マイクロホンにおいて，経路差 $d\sin(\theta)$ の分だけ，伝搬遅延時間が異なる信号として受信される．ビームフォーマでは，ある方向 θ_0 から到来する信号に関する伝搬遅延を補償するように，$\delta = d\sin(\theta_0)/c$（$c$ は音速）だけ，一方のマイクロホン信号を遅延させる．その出力信号を，他方のマイクロホン信号と加算または減算する．

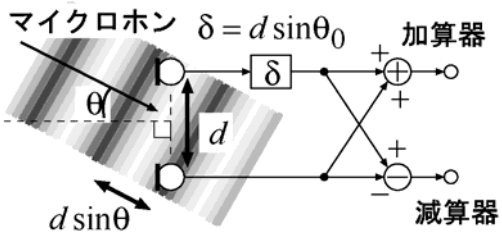
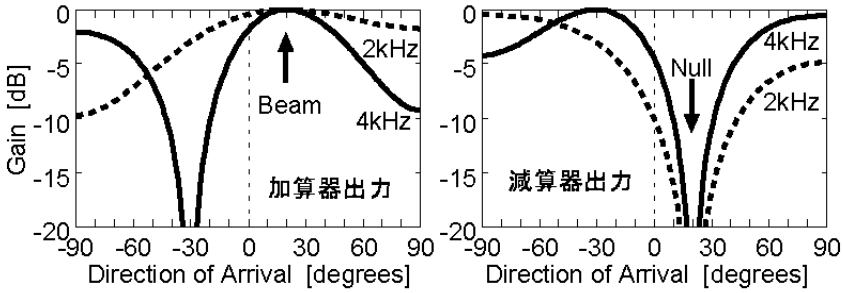


図 2・1　ビームフォーミングの原理

加算器の入力では，方向 θ_0 から到来する信号の位相が一致する．したがって，加算器出力において，方向 θ_0 から到来した信号は強調される．一方，θ_0 以外の方向から到来した信号は，互いに位相が一致しないため，θ_0 から到来した信号ほど強調されることはない．その結果，加算器出力を用いるビームフォーマは，θ_0 にビーム（Beam：特に感度の高い方向，ローブ）を有する指向性を形成する．

対照的に，減算器では，方向 θ_0 から到来する信号が完全にキャンセルされる．したがって，減算器出力を用いるビームフォーマは，**図 2･2** のように，θ_0 にヌル（Null：特に感度の低い方向）を有する指向性を形成する．



（マイクロホン間隔は 5 cm）

図 2・2　ビームフォーミングによる指向性の例

このように，遅延と加算のみを行うビームフォーマを，遅延和ビームフォーマ（Delay-and-Sum Beamformer）という．また，ビーム方向の制御をビームステアリング，ヌル方向の制御をヌルステアリングという．

ビームフォーミングは，両者を含む指向性制御技術一般を意味する．ビームを目標信号（所望信号）方向にステアリングし，ヌルを不要信号（妨害信号）方向にステアリングすることにより，目標信号を強調し，不要信号を抑圧することができる．

これまで最も簡単なビームフォーマである，マイクロホン 2 個と遅延器と加減算のみによる構成を用いて説明してきたが，当然，マイクロホン数を多くし，遅延だけではなく一般的なフィルタを用いた方が高性能になる（**図 2･3**）．多数のマイクロホンによって，空間的な自由度が高まり，鋭い指向性を得やすくなる．

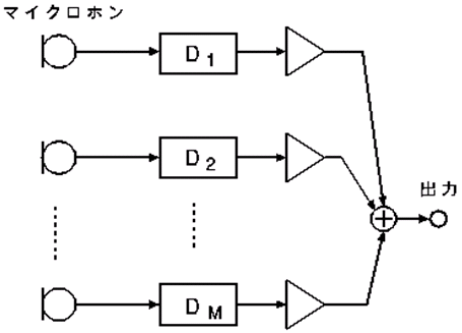


図 2・3　遅延和ビームフォーマの構成

また，フィルタの採用により，周波数と指向性の関係を変化させることができる．フィルタを用いた構成をフィルタアンドサムビームフォーマ（Filter-and-Sum Beamformer）という．その直接型構成を**図 2･4** に示す．

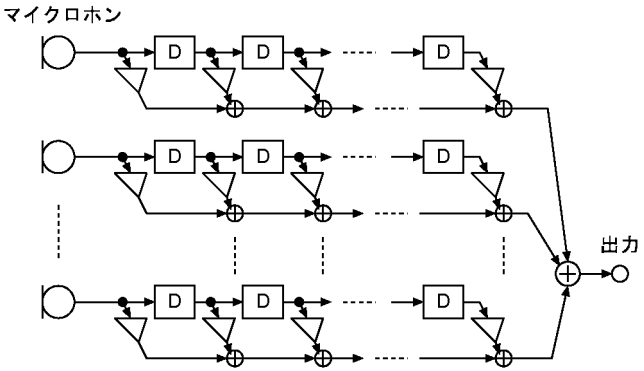


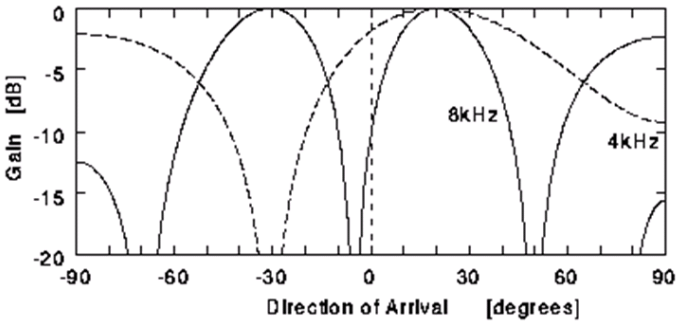
図 2・4　フィルタアンドサムビームフォーマ

2-1-2 音響ビームフォーミングの基本的問題

音響ビームフォーミングには以下のような問題がある．ビームフォーミングはアンテナやソナーにおいて古くから研究されているが，それらと共通の問題に加えて，音響（オーディオ）信号の周波数範囲（比帯域）は，Hi-Fi 用で 20 Hz から 20 kHz と 1000 倍，電話用でも 300 Hz から 4 kHz と 10 倍以上あることに由来する問題がある．

(1)　空間エリアシング

空間方向のサンプリング定理を満たすためには，マイクロホン間隔 d を最高周波数成分の半波長以下としなければならない．この条件が満たされない場合には，**図 2・5** における 8 kHz のように，グレーティングローブ（Grating Lobe）と呼ばれる空間的なエリアシングが生じる [3]．図 2･5 の 8 kHz では意図的にビームを向けた 20 度だけでなく－30 度にグレーティングローブは発生し感度が高くなっている．ただし，この空間領域のエリアシンクは，周波数領域でのエリアシングのようにノイズとして聞こえるわけではない．指向性合成において，積極的にグレーティングローブを利用することもできる．

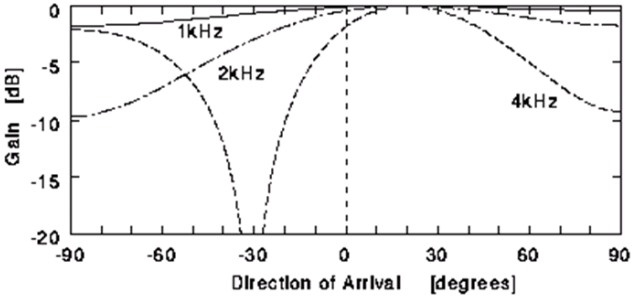


（マイクロホンは 2 個，間隔は 5 cm，ビームの方向は 20 度）

図 2・5　空間エリアシングの例

(2) 指向性の周波数依存性

ある周波数の指向性は,波長に対するマイクロホン間隔 d の比率によって決定されるため,周波数依存性がある [3]. **図 2･6**, 図 2･5 においても，指向性が周波数によって大きく異なることが分かる．音響信号への応用では，指向性の周波数依存性は小さいことが望ましい．



（遅延和ビームフォーマの場合．マイクロホン数 2 個，間隔 5 cm，ビームの方向は 20 度）

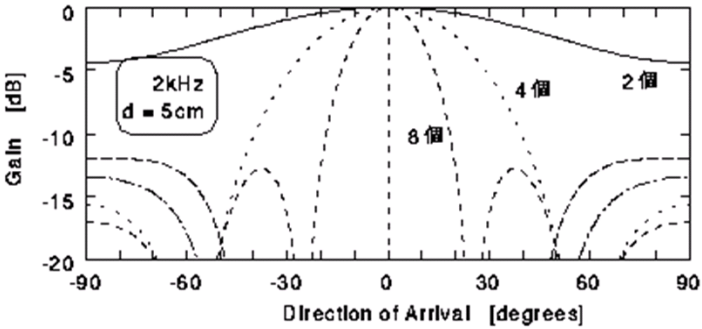
図 2・6 周波数と指向性の関係

(3) 低域指向性 対 アレーの大きさ

鋭いビームを形成するためには，マイクロホンアレー全体のサイズを波長に対して十分大きくしなければならない．高周波に対しては鋭い指向性が比較的容易に得られるが，低周波に対しては鋭い指向性を得ることは難しい [3]．300 Hz でも波長は 1 m 以上もある．

(4) マイクロホン数

マイクロホン数が多いほど，空間的な自由度が増えるため，**図 2･7** に示すように指向性を鋭くすることが可能になる [3].マイクロホン自身は安価であり,A/D(アナログ／ディジタル)変換器も高価ではないが，配線コストは安価とはいえない．一般的にはマイクロホン数は少ない方が望ましい．



（マイクロホン間隔は 5 cm，ビームの方向は 0 度）

図 2・7 遅延和ビームフォーマにおけるマイク数と指向性の関係

2-1-3 固定ビームフォーミング

信号到来方向など事前の知識のみから設計されるビームフォーミングを，後述の適応ビームフォーミングに対比して，固定ビームフォーミング（Fixed Beamforming）と呼ぶ．目標信号の到来方向のみが既知であり，不要信号到来方向は不明であるような応用は多い．このような状況では，目標信号到来方向に鋭いビームを向け，それ以外の方向に対する感度が低いような指向性を形成するべきである．

このような指向性は空間的なバンドパスフィルタであり，ディジタルフィルタの特性近似理論と同様の理論が適用できる [4]．ビームやヌルの方向と周波数応答に応じた多次元フィルタの設計法が提案されている．固定ビームフォーミングでは，限界などは理論的に明らかであるので，状況に即した設計手法が求められる．

(1) マイクロホン数削減技術

高周波における空間エリアシングを防止し，かつ，低周波において鋭い指向性を得ようとすると，マイクロホン数は膨大になる．この対策として，帯域分割に基づく手法が提案されている．高周波については中央付近に密に配置したマイクロホンのみを用い，低周波については，広範囲に粗く配置したマイクロホンを用いる [5]．中心から離れるに従って，マイクロホン間隔が粗になる．このマイクロホン配置方法はネスティング（Nesting：入れ子）と呼ばれる．二つの帯域に分割した場合の例を**図 2･8** に示す．高域，低域とも五つのマイクロホンを用いている．

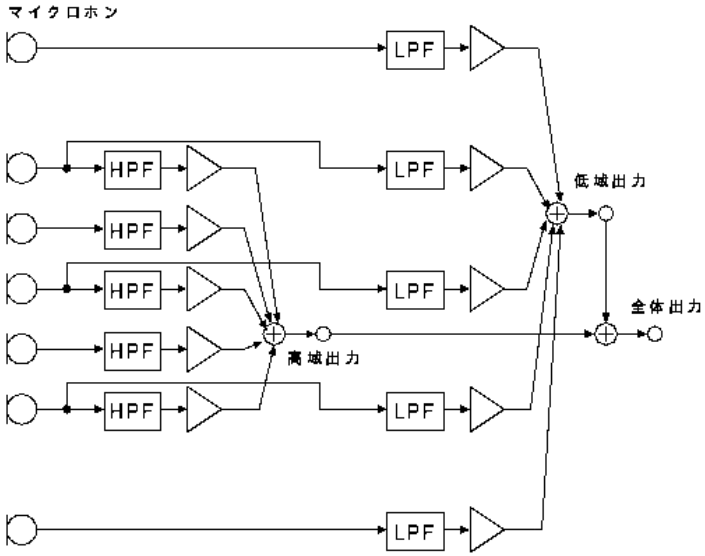


図 2･8　ネスティングを行ったマイクロホンアレーの構成例

(2) 指向性の周波数依存性低減

指向性の周波数依存性を減少させる設計法は，多数提案されている．直線配置のアレーに

対しては，時間周波数－空間周波数の 2 次元空間において，ファン型の通過域特性をもつフィルタ（ファンフィルタ）を設計することになる [6]．平面上の方形格子にマイクロホンを配置したアレーに対しては，時間周波数－空間周波数－空間周波数の 3 次元空間において，円錐型の通過域をもつフィルタ（コーンフィルタ）を設計する．ネスティングを行った場合の設計法も提案されている [7)–9)]．

(3) 伝搬モデル誤差の影響

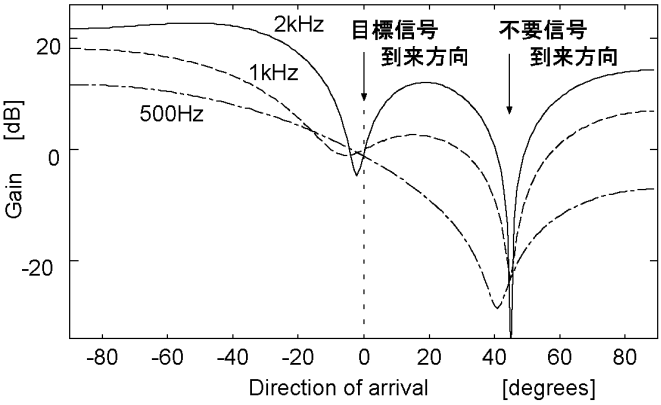
伝搬モデルが仮定と異なる場合，設計どおりの特性が得られなくなる．ほとんどの指向性制御アルゴリズムは，マイクロホン配置，平面波近似あるいは点音源近似，無反射仮定など，音源からマイクロホンまでの伝搬に，特定のモデルを用いている．実際には，室内音場（残響，反射，回折），マイクロホン配置の誤差，マイクロホン設置機構の干渉（反射，回折），マイクロホン相互の特性誤差（指向性，周波数特性，感度）など，モデルで考慮しなかった現象が生じる．

ヌルでは，誤差の影響がビームより深刻である．ヌルを形成するためには，位相と振幅が完全に一致した信号を減算しなければならない．位相や振幅に誤差がある場合には，深いヌルを形成することは困難である．また，気温や気圧の変化により音速が仮定と異なると，ヌルが設計どおりに形成されない．特性誤差を少なくするためには，シリコンマイクのような相対誤差の少ないマイクロホンを採用するか，マイクロホンの選別が必要である．特性の経年変化を考えると，オンラインのキャリブレーションが望ましいが，周波数と指向性の多次元キャリブレーションは容易でない [10)]．

2-1-4 適応ビームフォーミング

適応的に指向性を形成する信号処理を，適応ビームフォーミングという．適応ビームフォーミングでは，ビームステアリングとヌルステアリングを同時に行うことができる．しかし，音響ビームフォーミングでは，目標信号の方向を別途推定してビームステアリングは固定ビームフォーミングで行い，ヌルステアリングのみに適応信号処理（ここでは最適化を行う処理）を利用する場合が多い．適応ビームフォーミングでは，適応動作によって，誤差がある場合でも深いヌルが形成できる．そのため，一般的には，適応ビームフォーマの不要信号除去能力は固定ビームフォーミングより高い．ただし適応動作によって逆に目標信号を劣化させてしまう場合もある．

適応ビームフォーマにより得られた指向性の例を**図 2･9** に示す．目標信号の方向では，いずれの周波数においてもゲインはほぼ一定であるのに対し，不要信号方向には深いヌルが形成されている．指向性は，周波数ごとに全く異なるが，不要信号除去／目標信号抽出という目的は達成している．信号が存在しない方向への特性は全く関知していない．その分の自由度を，適応アルゴリズムを用いて不要信号除去と目標信号の保持に振り向けたことにより高い不要信号除去性能が得られているともいえる．



（適応ビームフォーマとしては，Griffiths-Jim ビームフォーマを用いた）

図 2・9　適応ビームフォーマにより得られた指向性の例

(1)　Frost ビームフォーマ

ここでは，代表的な適応ビームフォーマとして，Frost ビームフォーマ [11] と，Griffiths-Jim ビームフォーマ（GJBF）[12] のみを紹介する．Frost ビームフォーマ，GJBF とも線形拘束付最小分散型（Linear Constrained Minimum Variance : LCMV）ビームフォーマの一種である．ある規定した方向に対する周波数特性を拘束し，その拘束のもとで適応信号処理により出力信号を最小化している．Frost ビームフォーマでは，各マイクロホンについたフィルタ係数を拘束された空間に射影しながら更新する．

Frost ビームフォーマにおける拘束条件は線形方程式であるが，これは不等式であってもよい．拘束条件の緩和により，不要音除去性能が改善する．目標信号への周波数特性が乱れるが，主観的には大きな問題にはならない [13]．

(2)　Griffiths-Jim ビームフォーマ（一般化サイドローブキャンセラ）

GJBF は一般化サイドローブキャンセラ（Generalized Sidelobe Canceller : GSC）とも呼ばれる．GJBF は Frost ビームフォーマと等価であるが，固定ビームフォーマとブロッキング行列によって，フィルタ係数空間を，拘束された空間と，拘束と関係なく適応を行う空間とに分解していることが特徴である．これにより拘束された空間への射影演算が不要になり，演算量が削減される．Griffiths-Jim ビームフォーマ（GJBF）の構成を**図 2・10** に示す．GJBF は，固定ビームフォーマ（Fixed Beamformer : FBF），多入力キャンセラ（Multiple-input Canceller : MC），ブロッキング行列（Blocking Matrix : BM）から構成される．固定ビームフォーマは，規定方向から到来する目標信号（規定方向信号）を通過させ，規定方向以外から到来した信号を減衰させるような指向性を形成する．

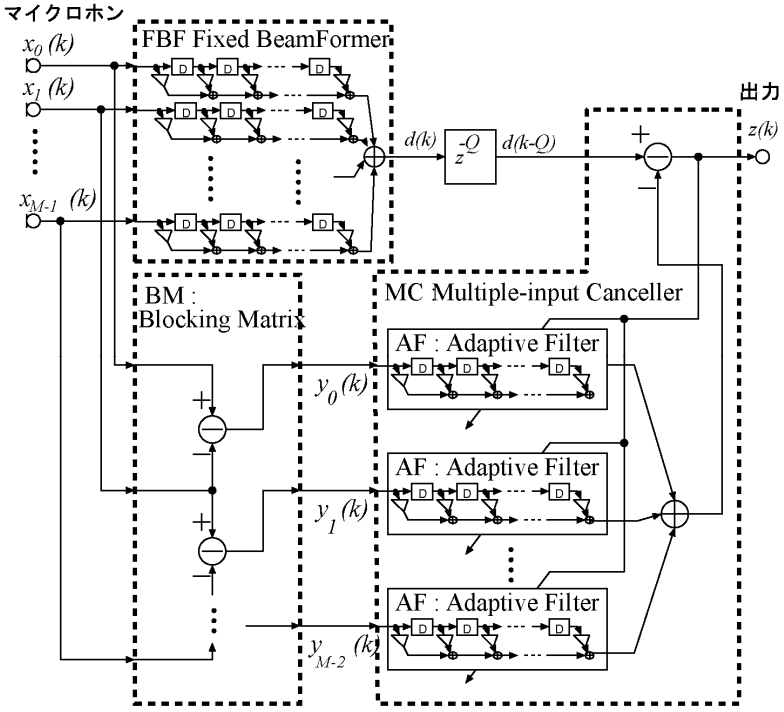


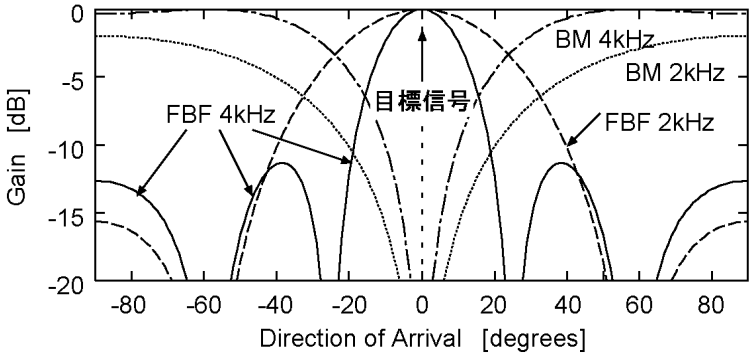
図 2・10　Griffiths-Jim ビームフォーマ（一般化サイドローブキャンセラ：GSC）の構成例

ブロッキング行列では，マイクロホン信号 $x_m(k)$（$m = 0, 1, \ldots M-1$）を $y_m(k)$（$m = 0, 1, \ldots M-2$）に変換する．$y_m(k)$は，規定方向信号成分が遮断され，かつ空間的に一時独立な（従属でない）信号である．その変換は，図 2･10 のように簡単な構成で実現できる．この変換は，規定方向信号が各マイクロホンにおいて同位相同振幅で得られることを利用しており，単に各マイクロホン信号の差をとることにより規定方向信号，すなわち目標信号のみが遮断される．規定方向が正面でない場合は遅延と増幅により，規定方向の信号が同位相同振幅で得られるように調整すればよい．ブロッキング行列という名は，目標信号を通過させずに（ブロッキング），ベクトル $x_m(k)$をベクトル $y_m(k)$に変換する行列であることに由来する．

多入力キャンセラは，複数の適応フィルタから構成される．これらの適応フィルタは，固定ビームフォーマ出力を遅延した信号 $d(k-Q)$から，ブロッキング行列の出力信号 $y_m(k)$に相関がある成分を除去する．$y_m(k)$は規定方向以外から到来した信号すなわち不要信号を含んでいるので，差信号 $z(k)$では，周囲騒音など不要信号が除去されている．また，$y_m(k)$は規定方向信号を含まないので, 規定方向信号に対する感度は影響を受けない. したがって, $z(k)$には，固定ビームフォーマを通過した目標信号がそのまま得られる．

GJBF の構成は，複数の参照信号を有する適応ノイズキャンセラにおいて，ブロッキング行列を適応フィルタの前処理として設けたものとも解釈できる．参照信号である雑音源信号

を得るために，適応ノイズキャンセラでは雑音近くにマイクロホンを設置するが，GJBF ではブロッキング行列の指向性を用いて雑音源信号に相関がある信号を抽出している．



（マイクロホン数 4 個，間隔 5 cm）

図 2・11　FBF と BM の指向性の例

ブロッキング行列出力において，目標信号が存在しなければ，いくら MC で適応キャンセルを行っても，目標信号に影響を与えることなく，不要信号を除去することができるはずである．しかし実際にはマイク間の相対誤差や室内の反響などによりブロッキング行列の出力には，目標信号が大きくもれこみ，不要信号とともに，目標信号もキャンセルされる．この問題に対処するビームフォーミングアルゴリズムはロバスト適応ビームフォーミング（Robust Adaptive Beamforming）と呼ばれる．ブロッキング行列を適応フィルタによって構成するなどの手法が提案されている [14), 15)].

適応ビームフォーマにおいても，帯域分割やアイゲンスペースなど，相関を低減した空間に変換しての処理は有効である．変換により自己相関が低減されるため，適応アルゴリズムの収束速度が改善される場合が多い [15), 16)].

■参考文献

1) M. Brandstein and D Ward eds., “Microphone Arrays: Signal Processing Techniques and Applications,” Springer, 2001.
2) J. Benesty, J. Chen, and Y. Huang, “Microphone Array Signal Processing,” Springer, 2008.
3) D. H. Johnson and D. E. Dudgeon, “Array Signal Processing －Concepts and Techniques－,” Prentice Hall, 1993.
4) C.L. Dolph, “A Current Distribution for Broadside Arrays Which Optimizes the Relationship Between Beamwidth and Side-Lobe Level,” Proc. IRE and Electrons, June 1946.
5) J.L. Flanagan, D.A. Berkley, G.W. Elko, J.E. West, and M.M. Sondhi, “Autodirective Microphone Systems,” Acustica, 73, pp. 58-71, Feb.1991.
6) 西川 清, “ビームフォーミングの 2 次元領域解析,” 信学論, vol.J77-A, no.9, pp.1304-1306, Sep. 1994.
7) M.M. Goodwin and G.W. Elko: “Constant Beamwidth Beamforming,” IEEE ICASSP'93, vol.I, pp.169-172, Apr. 1993.
8) T. Taniguchi, “Broadband Frequency Invariant Beamforming Method with Low Computational Cost,” IEEE ICASSP'98, vol.4, pp.2029-2032, May 1998.

9) R. Kennedy, D. Ward, and T. Abhayapala, “Nearfield Beamforming Using Nearfield/Farfield Reciprocity,” IEEE ICASSP'97, vol.5, pp.3741-3744, Apr. 1997.
10) M. Buck, T. Haulick, and H.-J. Pfleiderer, “Microphone Calibration for Multi-Channel Signal Processing,” in “Speech and Audio Processing in Adverse Environments,” ed. by E. Haensler, G. Schmidt, Springer, 2008.
11) O.L. Frost, “An algorithm for linearly constrained adaptive array processing,” Proceedings of IEEE, vol.60, no.8, pp.926-934, 1972.
12) L. J. Griffiths and C. W. Jim, “An Alternative Approach to Linear Constrained Adaptive Beamforming,” IEEE Trans. AP, vol.AP-30, no.1, pp.27-34, Jan. 1982.
13) Y. Kaneda and J. Ohga, “Adaptive Microphone-Array System for Noise Reduction,” IEEE Trans. ASSP, vol.34, no.6, pp.1391-1400, Dec. 1986.
14) O. Hoshuyama, A. Sugiyama, and A. Hirano, “A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters,” IEEE Trans. SP, vol.47, no.10, pp.2677-2684, 1999.
15) W. Herbordt, “Sound Capture for Human/Machine Interfaces - Practical Aspects of Microphone Array Signal Processing,” Springer, March 2005.
16) J. Mayer and G. W. Elko, “A Highly Scalable Spherical Microphone Array Based on an Orthonormal Decomposition of the Soundfield,” IEEE ICASSP2002, pp.1781-1784, 2002.

■2 群 － 6 編 － 2 章

2-2 独立成分分析に基づくブラインド音源分離

（執筆者：牧野昭二） ［2011 年 11 月受領］

独立成分分析に基づくブラインド音源分離は，複数音源が統計的に互いに独立であるという仮定のみを用い，分離信号が互いに独立となるようなフィルタを求める手法である [1), 2), 3), 4)]．この手法は，音源の種類や空間的位置の知識，目的音または妨害音区間の切り出し，更に，混合条件などの情報を原理的に必要とせず，音源信号の調波構造などの仮定も用いない．

2-2-1 信号モデルと分離システム

(1) 観測信号と分離信号

いくつかのマイクロホンを空間の異なる位置に配置すれば（マイクロホンアレー），音源信号は異なる時間差とレベル差でマイクロホンに混入する．音源信号が音であり混合系が部屋である環境では，マイクロホンで収音された混合信号は残響の影響を受ける．したがって，M 個のマイクロホンで収音された N 個の混合信号は（$M \geqq N$）

$$x_j(t) = \sum_{i=1}^{N} \sum_{p=1}^{P} h_{ji}(p) s_i(t-p+1) \qquad (j = 1,\ldots,M) \qquad (2・1)$$

とモデル化できる．ここで，s_i は音源 i からの音源信号，x_j はマイクロホン j で収音された混合信号，h_{ji} は音源 i からマイクロホン j への P タップのインパルス応答である．
分離システムは，Q タップの分離フィルタ w_{ij} を推定し，分離信号

$$y_i(t) = \sum_{j=1}^{M} \sum_{q=1}^{Q} w_{ij}(q) x_j(t-q+1) \qquad (i = 1,\ldots,N) \qquad (2・2)$$

を得る．分離フィルタは，分離信号が統計的に互いに独立になるように求める．以降，2 入力 2 出力の問題，つまり $N = M = 2$ で説明する（**図 2・12**）．

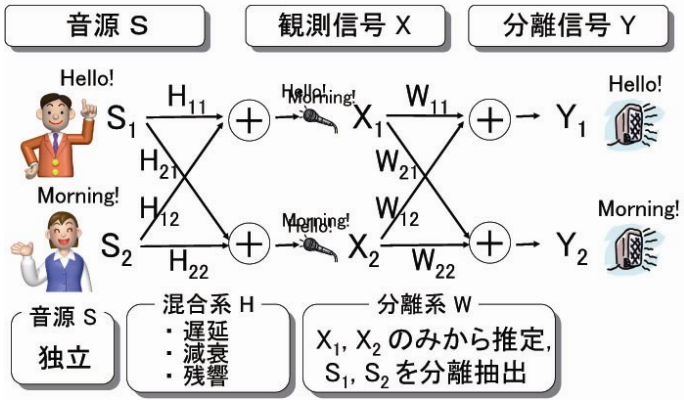


図 2・12　ブラインド音源分離モデル

(2) ブラインド音源分離システム

各信号を波形で見てみよう（**図 2･13**）．音源信号 s_1, s_2 は互いに独立であると仮定する．この仮定は，実環境の音源信号については，通常，成り立つ．混合信号を収音するマイクロホンを 2 本用いると，観測信号 x_1, x_2 には相関がある．この相関のある観測信号を入力として，出力 y_1, y_2 を互いに独立とする分離フィルタ w_{ij} を逐次的に学習し，出力 y_1, y_2 を分離・抽出する．この操作により，音源信号 s_1, s_2 の推定値が分離信号 y_1, y_2 に得られる．

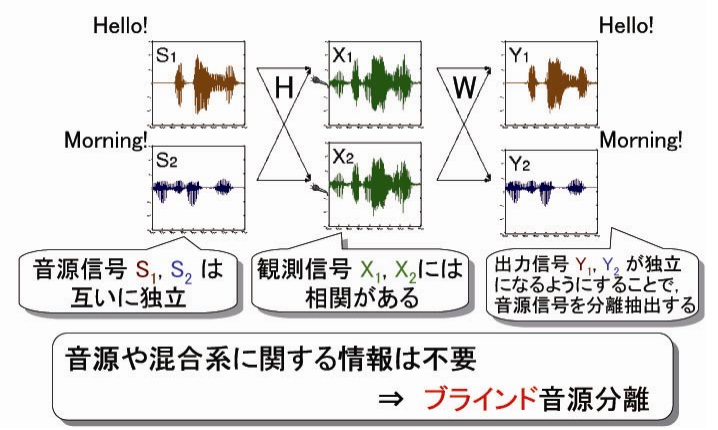


図 2･13　波形で見たブラインド音源分離

2-2-2 混合モデルと分離手法

(1) 瞬時混合と畳込み混合

これまで説明してきたように，部屋の中で音を分離する場合には，混合系 h_{ji} は P が数千タップに及ぶ FIR フィルタになる．この問題は畳込み混合の問題と呼ばれる．これに対して，混合系 h_{ji} で $P = 1$ である場合，すなわち，遅延や残響がなく，例えば，ミキサーを使って音をミキシングしたような場合には，瞬時混合の問題と呼ばれる．

(2) 時間領域手法と周波数領域手法

畳込み混合の問題を解くために，時間領域手法は，混合系のインパルス応答 h_{ji} を FIR フィルタで表し，分離フィルタを時間領域で推定する．一方，周波数領域手法は，時間領域の畳込み混合を，周波数領域の複数の瞬時混合に変換して解く．

(3) 畳込み混合に対する周波数領域手法

式(2･1)に短時間フーリエ変換を施し，ベクトル表記すれば，周波数領域の時間系列信号が得られる．

$$\boldsymbol{x}(\omega, k) = \boldsymbol{H}(\omega)\, \boldsymbol{s}(\omega, k) \qquad (2 \cdot 3)$$

ここで，ω は周波数，k は時間，$\boldsymbol{s}(\omega, k) = [S_1(\omega, k), S_2(\omega, k)]^T$ は音源信号ベクトル，$\boldsymbol{x}(\omega, k) = [X_1(\omega, k), X_2(\omega, k)]^T$ は観測信号ベクトル，$\boldsymbol{H}(\omega)$ は周波数 ω における(2×2)混合行列である．分離信号

は各周波数 ω で

$$\boldsymbol{y}(\omega, k) = \boldsymbol{W}(\omega)\,\boldsymbol{x}(\omega, k) \qquad (2・4)$$

と表される．ここで，$\boldsymbol{y}(\omega, k) = [Y_1(\omega, k), Y_2(\omega, k)]^T$ は分離信号ベクトル，$\boldsymbol{W}(\omega)$は周波数 ω における(2×2)分離行列であり，分離信号 $Y_1(\omega, k)$, $Y_2(\omega, k)$が互いに独立になるように求める．この計算は各周波数でそれぞれ行われる．

2-2-3 独立成分分析

独立成分分析は統計的な手法で，その理論には，信号どうしの統計的独立性という，統計理論において最も一般的な特徴が利用されている．独立成分分析は，観測信号 X_j のみから，分離信号 Y_i が互いに独立となるような線形な分離行列 $\boldsymbol{W}(\omega)$と分離信号 Y_i の両方を推定する手法である．

「独立」という概念は「無相関」の概念より強い．すなわち，相関が 2 次の統計量に基づくものであるのに対して，独立は高次の統計量に基づく．簡単にいえば，独立とは，片方の信号がもう一方の信号に関する情報をもっていないということである．独立な成分は，高次統計量に基づく非線形相関除去により求めることができる（2 次統計量と音源信号の非定常性や非白色性に基づく，非定常相関除去，非白色相関除去については，文献参照）[1), 2)].

独立成分分析には，相互情報量の最小化，非ガウス性の最大化，ゆう度の最大化，の三つの理論があるが，面白いことに，上記三つの解は同一である [5), 6), 7)].

(1) 教師なし学習

まず, 初期状態にある分離行列 $\boldsymbol{W}(\omega)$を用いて式(2・4)の操作により分離信号 Y_1, Y_2を求める．次に，分離行列 $\boldsymbol{W}(\omega)$を変化させ，分離信号 Y_1, Y_2 間の相互情報量を最小化する，非ガウス性を最大化する，あるいは，ゆう度を最大化する，分離行列 $\boldsymbol{W}(\omega)$を求める．この更新を繰り返す，いわゆる教師なし学習を経て，システムは互いに独立な分離信号を生成する．この操作は，勾配法などにより実現できる．

(2) 高次統計量に基づく手法

分離行列 $\boldsymbol{W}(\omega)$を求めるために, Kullback-Leibler Divergence の最小化に基づくアルゴリズムが提案されている [5), 6)]．安定で収束速度の速いアルゴリズムとして，ナチュラルグラジェントに基づくアルゴリズムが甘利によって提案された [7)]．ナチュラルグラジェントを用いれば，最適な分離行列 $\boldsymbol{W}(\omega)$は次のような逐次勾配法によって

$$W_{i+1} = W_i + \mu \begin{bmatrix} 1-\langle \phi(Y_1)Y_1^* \rangle & \langle \phi(Y_1)Y_2^* \rangle \\ \langle \phi(Y_2)Y_1^* \rangle & 1-\langle \phi(Y_2)Y_2^* \rangle \end{bmatrix} W_i \qquad (2・5)$$

と表される．ここで，ϕ は非線形関数，＊は複素共役，〈・〉は平均操作，i は繰り返しの i－番目，μ はステップサイズを表す．

(3) スケーリングとパーミュテーション

周波数領域手法には，各周波数が個別に取り扱われるために生じるパーミュテーションの問題がある．これは，分離信号の各周波数成分はそれぞれの周波数で別々の順番で現れると

いう問題である．周波数領域手法のパーミュテーションの解法として，音源の方向情報と分離信号の相関を利用した方法がある [8]．

ブラインド音源分離では，スケーリングの問題も大きな問題である．分離信号の各周波数成分はそれぞれの周波数で別々のゲインで得られる．各周波数におけるスケーリングの任意性は，分離信号の畳込みの任意性，すなわち，フィルタリングの任意性となって現れる．このことは，独立な信号をフィルタリングしたものもまた独立であるという事実を反映している．スケーリングの解法として，Minimal Distortion Principle に基づく方法がある [9]．

2-2-4 ブラインド音源分離の音響信号処理からの解釈

独立成分分析に基づくブラインド音源分離は，適応ビームフォーマと呼ばれるマイクロホンアレーと同じ動作原理を実現している [1), 2)]．

マイクロホンが 2 本の場合，適応ビームフォーマは妨害音方向に適応的な空間的死角を一つ形成し，目的音を抽出する．ブラインド音源分離も適応ビームフォーマと同様に，妨害音方向に適応的な死角を一つ形成し，目的音を抽出する（**図 2･14**）．

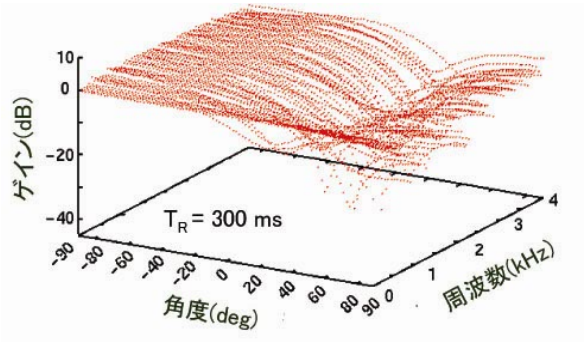


図 2・14　ブラインド音源分離の指向性パターン（TR：**残響時間**）

適応ビームフォーマと違って，ブラインド音源分離には，マイクロホンの位置や音源の情報などは不要である．適応ビームフォーマでは，目的音の位置情報を拘束条件としながら，目的音が無く妨害音のみが鳴っている時間を検出して，その時だけ出力誤差に対する 2 乗誤差最小化の規範により適応動作を行う．適応ビームフォーマにおいて，出力誤差に対する 2 乗誤差最小化の規範は，特に目的音方向情報に誤差がある場合，妨害音のみが鳴っている時間の検出誤りに影響される．これに対して，ブラインド音源分離では，分離信号間の相関除去の規範により適応動作を行うため，目的音の位置情報や妨害音のみが鳴っている時間の検出が不要である．

以上のように考えれば，ブラインド音源分離は適応ビームフォーマの高機能版といえる．

■2 群 － 6 編 － 2 章

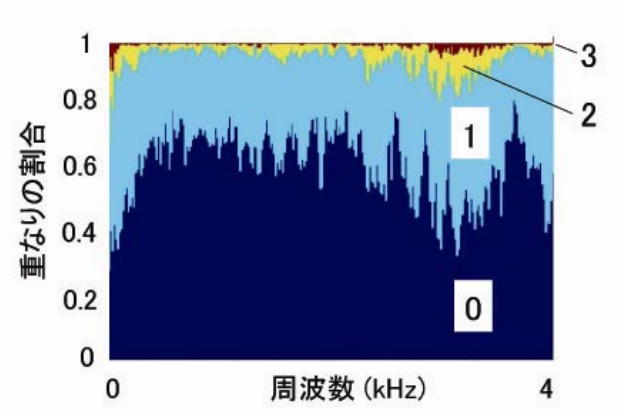
2-3 信号のスパース性を用いたブラインド音源分離

（執筆者：牧野昭二） ［2011 年 11 月受領］

前節の独立成分分析に基づくブラインド音源分離は，マイクロホン数 M ≧音源数 N の場合のみを扱い，マイクロホン数 M ＜ 音源数 N の場合へ適用することはできない．一方，信号のスパース性を用いた方法は，マイクロホン数 M ＜ 音源数 N の場合にも適用できる [10]．

2-3-1 音声信号のスパース性

信号がスパースであるとは，信号がほとんどの時間周波数において 0 であることを指す．信号のスパース性を仮定することで，複数の信号が同時に存在していても，各時間周波数ポイントで見れば互いに重なりあって観測される頻度は低いことを仮定できる．**図 2･15** は三つの音声信号の各時間周波数における重なりの一例を示したもので，約 6 割が信号なし（最大値から－20 dB 以下），約 3 割が 1 信号のみ，二つの信号が重なっているのは 1 割以下で，三つの信号が重なっているところはほとんどない．



紺色：信号なし（最大値から－20 dB 以下），水色：信号一つ，黄色：信号二つ，赤色：信号三つ

図 2・15　三つの音声信号の重なり

2-3-2 混合ベクトルの推定

スパース性の仮定は，時間領域より時間周波数領域でよりよく成立する（**図 2･16**）．時間周波数領域での観測信号 $X_j(\omega, k)$は

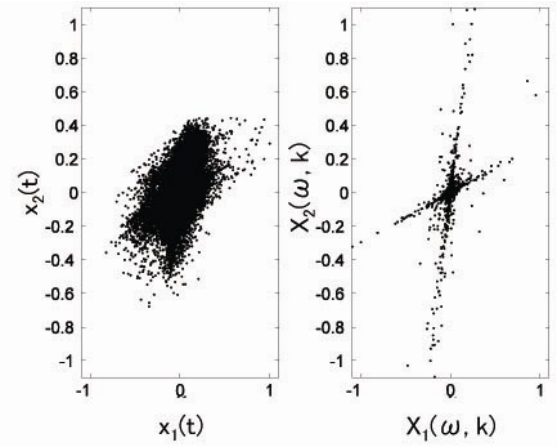
$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} H_{11}\, H_{12}\, H_{13} \\ H_{21}\, H_{22}\, H_{23} \end{bmatrix} \begin{bmatrix} S_1 \\ S_2 \\ S_3 \end{bmatrix} = \begin{bmatrix} H_{11} \\ H_{21} \end{bmatrix} S_1 + \begin{bmatrix} H_{12} \\ H_{22} \end{bmatrix} S_2 + \begin{bmatrix} H_{13} \\ H_{23} \end{bmatrix} S_3 \qquad (2 \cdot 6)$$

とモデル化できる．ここで $S_i(\omega, k)$（i = 1, …, N）は音原信号の短時間フーリエ変換の結果，$H_{ji}(\omega)$は音源 i からマイクロホン j への周波数応答である．

信号のスパース性から，各時間周波数(ω, k)において音源信号のうちの一つのみが支配的であると仮定する．例えば，$S_1 \neq 0, S_2 = 0, S_3 = 0$ とすれば式(2･6)は

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} H_{11} \\ H_{21} \end{bmatrix} S_1 = h_1 S_1 \qquad (2 \cdot 7)$$

となる．したがって，(X_1, X_2)をプロットすれば，混合ベクトル $\boldsymbol{h}_1$ に沿って分布する．この直線を求めれば，混合ベクトル $\boldsymbol{h}_1$ が求まる．2 音源の場合，二つの混合ベクトル $\boldsymbol{h}_1$, $\boldsymbol{h}_2$ に対応した 2 本の直線が得られる（図 2･16）．



二つの混合ベクトル $\boldsymbol{h}_1$, $\boldsymbol{h}_2$ に対応した 2 本の直線に沿って分布する．
左図：時間領域，右図：時間周波数領域

図 2･16　2 音源の場合の (X_1, X_2)の分布の例

音声信号の畳込み混合の場合には，X_1, X_2 は複素数となる．そのため，クラスタリングに用いる特徴量として，二つのマイクロホンで観測した各時間周波数領域信号の，振幅比や位相差（またはこれから推定される信号の到来方向）情報 [11)~13)] などが使われる．

クラスタリングの例を**図 2･17** に示す．二つのクラスタが形成されており，それぞれのクラスタが各音源信号 S_i に相当する．

2-3-3 信号の分離(各成分の振り分け)

(1)　バイナリマスク（ハードマスク）

それぞれのクラスタが個々の音源信号 S_i に相当するので，それぞれの信号に 0 or 1 のバイナリマスク（ハードマスク）を掛けることにより，それぞれのクラスタに属する時間周波数の観測信号を再構成して，それぞれの分離信号を得ることができる [11), 12)]．

(2)　L_1 ノルム最小化による信号の分離（ソフトマスク）

僅かではあるが信号が重なっている部分を，各成分に振り分けるために，ソフトマスクも

使われる．劣決定問題（マイクロホン数 M ＜音源数 N）であるため，前項で推定した混合ベクトル $\boldsymbol{h}_1$, $\boldsymbol{h}_2$ を用い，音源 S_i の L_1 ノルム和最小化という条件を付加して，分離信号を求める [13]．

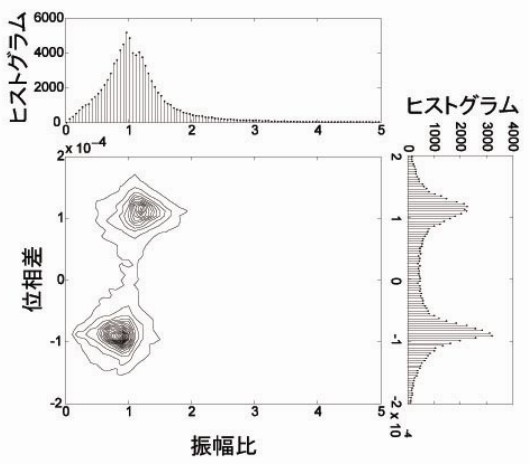


図 2・17　2 音源の場合の観測信号ベクトルのクラスタリングの例

■参考文献 (2-2, 2-3 節)

1) S. Makino, “Blind source separation of convolutive mixtures of speech,” in “Adaptive Signal Processing: Applications to Real-World Problems,” ed. by J. Benesty and Y. Huang, Springer, Berlin, Jan. 2003.
2) 牧野昭二, 荒木章子, 向井　良, 澤田　宏, “独立成分分析に基づくブラインド音源分離,” ディジタル信号処理シンポジウム, A3-2, Nov., 2003.
3) J. Benesty, S. Makino, and J. Chen, “Speech Enhancement,” Springer, Mar., 2005.
4) 澤田　宏, 荒木章子, 牧野昭二, “音源分離の最新動向,” 信学誌, vol.91, no.4, pp.292-296, Apr., 2008.
5) A. Hyvarinen, J. Karhunen, and E. Oja, “Independent Component Analysis,” John Wiley & Sons, 2001.
6) S. Haykin, “Unsupervised Adaptive Filtering,” John Wiley & Sons, 2000.
7) A. Cichocki and S. Amari, “Adaptive Blind Signal and Image Processing,” John Wiley & Sons, 2002.
8) 澤田　宏, 向井　良, 荒木章子, 牧野昭二, “多音源に対する周波数領域ブラインド音源分離,” AI チャレンジ研究会, SIG-Challenge-0522-3, pp.17-22, Oct., 2005.
9) K. Matsuoka and S. Nakashima, “Minimal distortion principle for blind source separation,” in “Proc. ICA,” pp.722-727, Dec., 2001.
10) S. Makino, Te-Won Lee, and H. Sawada, “Blind Speech Separation,” Springer, Sep., 2007.
11) M. Aoki, M. Okamoto, S. Aoki, H. Matsui, T. Sakurai, and Y. Kaneda, “Sound source segregation based on estimating incident angle of each frequency component of input signals acquired by multiple microphones,” Acoustical Sci. Technol., vol.22, no.2, pp.149-157, 2001.
12) O. Yilmaz and S. Rickard, “Blind separation of speech mixtures via time-frequency masking,” IEEE Trans. Signal Processing, vol.52, no.7, pp.1830-1847, July, 2004.
13) S. Makino, S. Araki, S. Winter, and H. Sawada, “Underdetermined blind source separation using acoustic arrays,” in “Handbook on Array Processing and Sensor Networks,” ed. by S. Haykin and K.J. Ray Liu, Wiley, Mar., 2008.