

# 4 正方形マイクロホンアレイによる音源分離技術

矢頭　隆*　森戸　誠　山田　圭*　小川 哲司**
* 沖電気工業（株）　** 早稲田大学

## 音源分離技術

音声認識技術の進展により，静粛環境ではかなり高精度な認識が実現できるようになった．しかし，実環境では話者の音声には周囲からのさまざまな雑音が混入し認識性能を著しく劣化させる．目的とする音声を背景音から分離することができれば，音声インタフェースを実用化する上できわめて有用である．

目的音をその他の雑音から分離抽出する技術は音源分離と呼ばれ，盛んに研究が行われている．雑音には，空調音のように比較的定常ではあるが音源が1方向に特定できない拡散性雑音と，音声や音楽のように指向性があって時間変動の大きい指向性雑音がある．両者は，その性質の違いから対策方法も異なる．一般に雑音抑圧技術と呼ばれるものも含め，以下に代表的な方法をいくつか紹介する．

拡散性雑音に対しては，スペクトルサブトラクション，ウィーナーフィルタ，コムフィルタなどの単一マイクロホンを用いた方式がよく知られている．

スペクトルサブトラクションは，雑音が定常であると仮定し，非音声区間の信号から雑音のパワースペクトルを推定し，推定した雑音成分を入力信号のパワースペクトルから引き去ることで雑音の低減を行う方法である．簡単な構成ながら定常雑音には非常に効果的であることから広く用いられている．一方，非定常雑音への対処や雑音下での音声区間検出が必要となる．

ウィーナーフィルタは，フィルタを通すことで得られる復元信号と原信号との差の二乗平均を最小にするフィルタである．周波数領域で，原信号を $S$，雑音が重畳した観測信号を $X$，雑音を $N$，復元信号を $Y$ とすると，その関係は，$Y=W\cdot X=W\cdot(S+N)$ となる．$W$ はウィーナーフィルタの伝達関数であり，$W=S_p/(S_p+N_p)$ で与えられる．ここで $S_p$ および $N_p$ は，原信号および雑音信号のパワースペクトルである．原信号 $S$ は直接観測できないため，$S_p=X_p-N_p$ として観測信号 $X$，あるいは1つ前の処理フレームの復元信号を用いて近似する．スペクトルサブトラクション同様，雑音の推定が必要である．

コムフィルタは，音声波形が周期的でありスペクトル上で調波構造を持つことを利用して音声成分を分離するものである．音声の基本周波数を推定し，雑音混じりの音声信号に対して基本周波数の高調波成分を強調する櫛形フィルタを作用させる．非定常な雑音にもある程度対応可能だが，周期性のない子音部には適用できないことや，雑音下でいかに精度よく基本周波数を推定するかが課題である．

これらの拡散性雑音に対する手法は，音声と雑音の性質の違いを利用して雑音の抑圧を行うものであり，目的とする話者以外の声，テレビ，ラジオの音など，音声そのものあるいは音声と類似した性質を持った妨害音に対しては有効に機能しない．そのような指向性雑音に対してはマイクロホンアレイを用いることが効果的である．マイクロホンアレイでは，各マイクロホンに到達する信号の時間差を利用して，特定の方向から到来する信号を強調あるいは抑圧することができる．代表的な方法として遅延和アレイと減算型アレイがある．

遅延和アレイは，時間差を持ってマイクロホンに到達する信号に対し，受音信号に時間差に相当する遅延を与えることで同相化し，それらを加算することにより，特定の方向から到来する信号のみを強調するものである．目的音方向に鋭い指向性を持たせるためには大規模のアレイが必要になる．

一方，減算型アレイは少数のマイクロホンで鋭い

## 4 正方形マイクロホンアレイによる音源分離技術

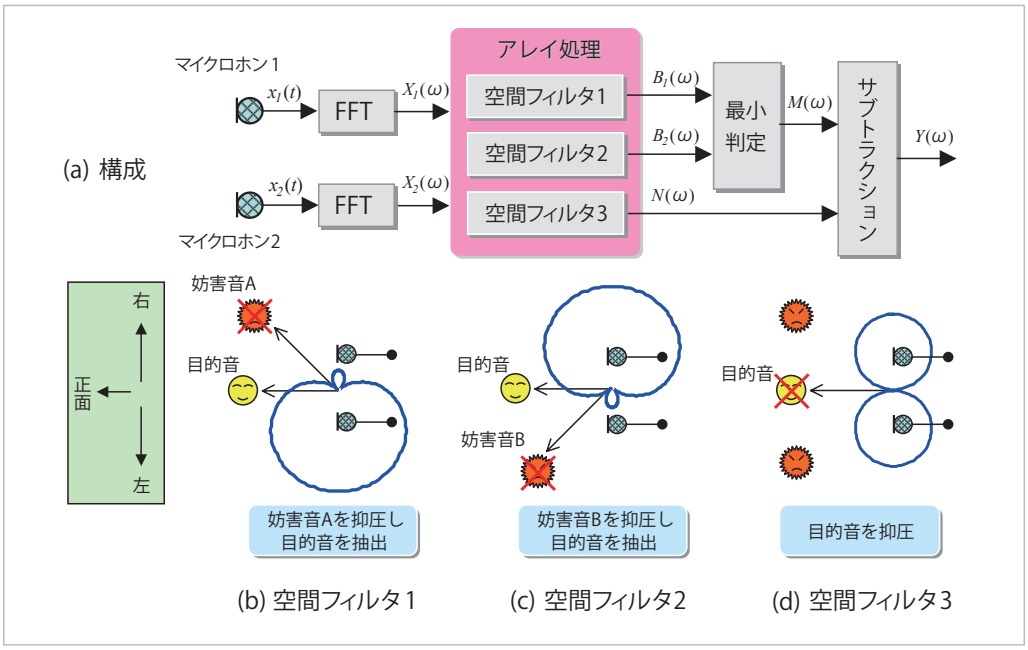


図-1　2チャネル音源分離

指向性を実現するため，指向性の谷すなわち死角を利用する．2つのマイクロホンの信号を特定の方向に対して同相化し減算すれば，互いの信号が相殺されて完全に抑圧される．減算型アレイでは，雑音の到来方向に死角を向けることで雑音を抑圧する．ただし，目的音の方向は運用によって限定することはできても，雑音の到来方向は一方に定まらないため，どのようにして雑音方向に死角を向けるかが問題となる．

同じく複数のマイクロホンを用いる方法として，音源の方位やマイクロホンの位置関係などの空間情報をまったく用いず，観測信号のみから音源信号を推定するブラインド音源分離がある．中でも独立成分分析（ICA：Independent Component Analysis）に基づく手法がよく用いられる．ICAでは，「音源は互いに独立である」という仮定のみを利用して，出力が互いに独立になるように分離フィルタを学習する．演算量は多いが，前記のように音源方向やマイクロホン配置の知識を使用しないため，マイクロホンの事前調整が不要というメリットがある．

音声インタフェースの操作端末に容易に実装可能な小型音声分離装置を開発するためには，少数のマイクロホンでコンパクトに実装可能であり，遅延が少なく，演算量も少ない音源分離技術が必要とされる．また，実環境においては指向性雑音，拡散性雑音いずれの雑音に対しても抑圧可能でなければならない．

本稿では，4個の無指向性マイクロホンを正方形の各頂点に配置した正方形マイクロホンアレイを用いた音源分離技術[1]を紹介する．4通りのマイクロホンペアによる減算型アレイ出力を用いた指向性雑音抑圧と，同じく減算型アレイ出力を用いたマルチチャネルウィーナーフィルタとシングルチャネルウィーナーフィルタの組合せにより拡散性雑音も同時に抑圧する．また，実際に本方式を搭載した音声分離小型モジュールを開発した[2]．このモジュールは，マイクロホンが縦横3cmの間隔で非常にコンパクトに配置されており，設置面積が限られる小型端末にも十分搭載可能であることを示した．

### 2チャネル音源分離

指向性雑音抑圧に関して，基本となる2チャネル音源分離処理[3]（図-1）を説明する．2チャネル音源分離では，2つのマイクロホンの入力信号に対して減算型のアレイ処理を施す．ここで減算型アレ



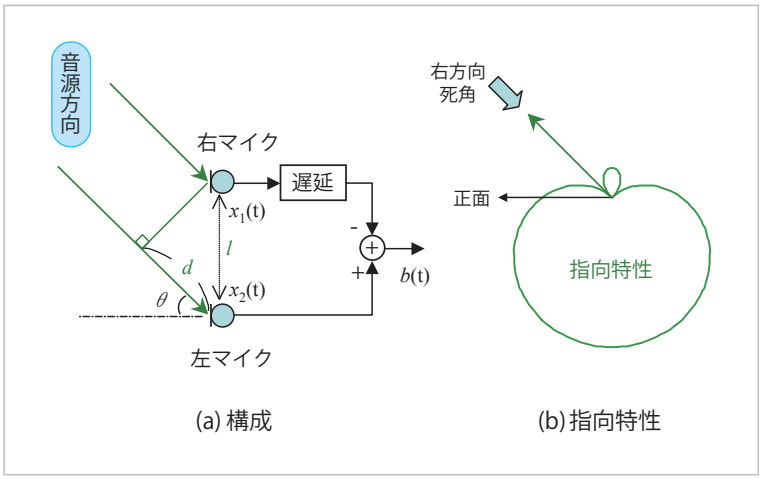


図 -2　減算型アレイの原理

イの原理（図 -2(a)）を説明する．角度 $\theta$ の方向から到来する平面波を距離 $l$ だけ離れて設置された左右 2 つのマイクロホンで受音することを考える．$\theta$ 方向から到来した音波は，まず音源に近いマイクロホン 1 に受音される．次に音波は距離 $d$ だけ進んでマイクロホン 2 に到達する．距離 $d$ は

$$d=l\sin\theta$$

と表される．したがって，マイクロホン 2 での受音信号 $x_2(t)$ はマイクロホン 1 での受音信号 $x_1(t)$ と比べて音波が距離 $d$ だけ進行するのに要する時間 $\tau$ だけ遅れた信号となっている．

$$x_2(t)=x_1(t-\tau)$$

$$\tau=d/c=l\sin\theta/c\quad(c：音速)$$

したがって $x_1(t)$ に遅延 $\tau$ を与え $x_2(t)$ から減算（逆位相で加算）すれば，

$$b(t)=x_2(t)-x_1(t-\tau)$$

信号同士が相殺され，角度 $\theta$ の方向に死角を持った指向性フィルタ（空間フィルタ）が形成される（図 -2(b)）．

このような時間軸上での空間フィルタ形成操作は，周波数領域でも同様に行うことができる．時間軸を $\tau$ だけ遅らせた信号のフーリエ変換は，もとの信号をフーリエ変換した結果に $e^{-j\omega\tau}$ を乗じたものになる．周波数領域の減算型アレイ処理は，$x_1(t)$ と $x_2(t)$ の短時間フーリエ変換 $X_1(\omega)$，$X_2(\omega)$ を用いて次のように表される．

$$B(\omega)=X_2(\omega)-e^{-j\omega\tau}X_1(\omega)$$

減算型アレイに与える遅延量 $\tau$ は，マイクロホン間隔 $l$ と方向 $\theta$ によって定まるが，離散信号における時間軸上の遅延操作は，$x(n-k)$（遅延量 $k$ は整数）のようにサンプリング周期単位に限定される．そのため形成できる死角方向に制約がある．一方，周波数領域では上式の $\tau$ に遅延時間を与えればよく，容易に所望の特性を得ることができる．

2 チャネル音源分離（図 -1(a)）では，2 つのマイクロホンの入力信号に対して周波数領域での減算型アレイ処理を施し 3 つの空間フィルタを形成する．空間フィルタ 1 は右方向に死角が設定されており（図 -1(b)），右方向から到来する妨害音を抑圧する．目的音は，ある利得を持って出力される．この出力を $B_1(\omega)$ とする．空間フィルタ 2 は左方向に死角が設定されており（図 -1(c)），左方向から到来する妨害音を抑圧する．空間フィルタ 1 と同様，目的音はある利得を持って出力される．この出力を $B_2(\omega)$ とする．空間フィルタ 3 は，正面方向に死角が設定され（図 -1(d)），目的音を抑圧する．この出力を $N(\omega)$ とする．空間フィルタ 1 の出力の振幅成分 $|B_1(\omega)|$ と空間フィルタ 2 の出力の振幅成分 $|B_2(\omega)|$ の小さいほうを選択する．

$$M(\omega)=\min[|B_1(\omega)|,|B_2(\omega)|]$$

右方向に妨害音音源が存在した場合，右方向に死角を持つ空間フィルタ 1 の出力 $B_1(\omega)$ は，妨害音が

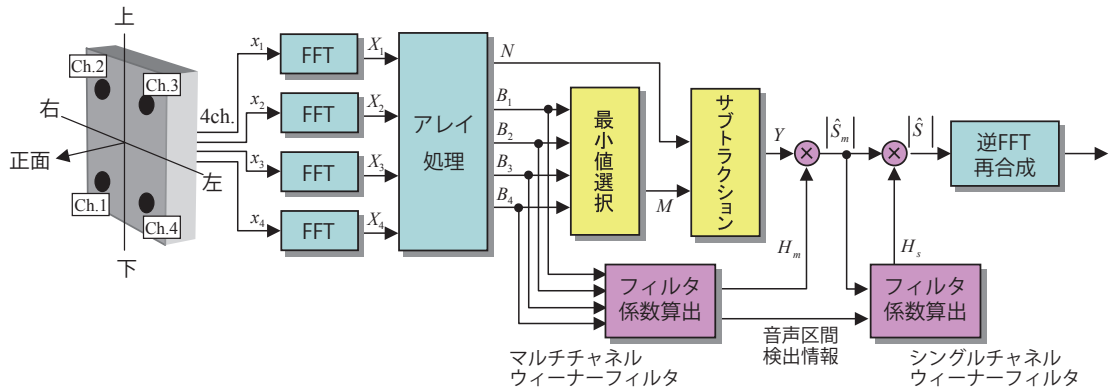


図 -3　正方形マイクロホンアレイによる音源分離

抑圧されて振幅が小さくなる．これに対し妨害音が存在しない方向に死角を持つ空間フィルタ 2 の出力 $B_2(\omega)$ には振幅に大きな変化はないと考えられる．逆に，左方向に妨害音源があれば $B_2(\omega)$ は小さくなるが $B_1(\omega)$ の変化は少ない．したがって最小値選択された $M(\omega)$ は，妨害音を抑圧した目的音候補成分となっている．最後に $M(\omega)$ と $N(\omega)$ によって以下のように帯域選択とスペクトルサブトラクションを行い出力を決定する．

$$Y(\omega)=\begin{cases}\sqrt{|M(\omega)|^2-\alpha|N(\omega)|^2} & if\ |M(\omega)|>\alpha|N(\omega)|\\ 0 & otherwise\end{cases}$$

ここで $\alpha$ は空間フィルタゲイン補正係数である．帯域選択は，信号に目的音の成分が含まれているかどうかを判定するために行う．$N(\omega)$ は目的音方向以外からの周囲雑音と考えられるから $N(\omega)$ が $M(\omega)$ より大きい場合は，そもそも目的音の成分が存在しない区間とみなして棄却する．$M(\omega)$ に目的音の成分があると判断されれば，サブトラクションを行って正面方向により鋭い指向性を向け目的音を分離する．

フィルタ 1 とフィルタ 2 の減算型アレイによる雑音抑圧は右か左かの粗いレベルであっても，後段のフィルタ 3 のサブトラクションによって目的音方向に鋭い指向性を形成するため，正面以外の指向性雑音に対し十分な抑圧効果が得られる．ここでは減算型アレイを雑音方向の抑圧だけではなく目的音方向の分離に用いていることに特徴がある．

簡単のため，ここでは 2 マイクでの構成を示したが左右方向だけでなく上下方向にもマイクを配置すれば，空間中の種々の方向からの指向性雑音に対応可能になる．

指向性による目的音強調という点では，ショットガンマイクロホンと呼ばれる超指向性マイクロホンがある．ショットガンマイクロホンは側面にスリットの入った円筒状の干渉管をマイクロホンユニットの先端に装着し，側面から入る音と干渉管の先端から入る音を干渉させ横方向からの音を抑圧する．干渉管には高度な設計と複雑な加工が必要で高価な上，20cm 程度の長さが必要である．一方，本方式は安価なマイクロホンと信号処理の組合せで実現でき，将来的に専用チップとして量産すれば低価格化が可能である．また実現できる指向性も干渉管による音響的な指向性形成に比べ，遥かに鋭いものとなる．

## 正方形マイクロホンアレイによる音源分離

実際の使用環境では指向性雑音だけが存在することはごく稀であり，指向性および拡散性雑音が混在して存在する．ここでは拡散性雑音も同時に抑圧する正方形マイクロホンアレイによる音源分離 [1)]（図 -3）について述べる．全体は指向性雑音抑圧部，拡散性雑音抑圧部，残留雑音抑圧部から構成される．



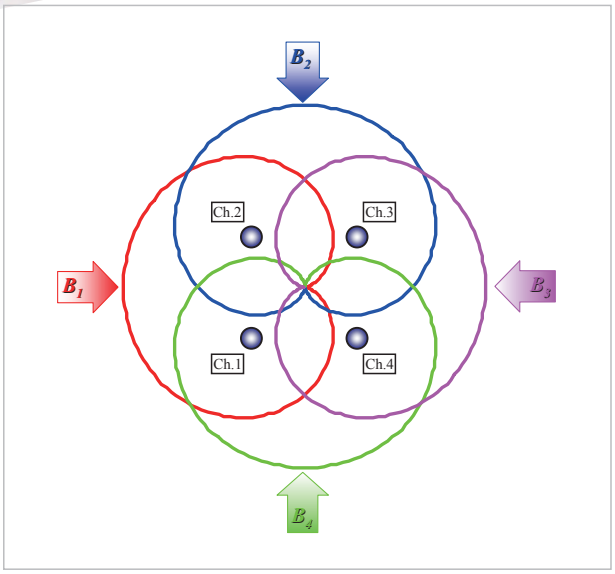


図 -4　空間フィルタ指向特性

入力には，平面上に 4 個の無指向性マイクロホンを正方形に配置した正方形マイクロホンアレイを用いる．目的音は正面方向から到来するものとする．

### ●指向性雑音抑圧

4 個のマイクロホンのうち，正方形各辺両端の 2 個ずつを組み合わせた 4 通りのペアを作る．それぞれのマイクロホンペアの減算型アレイによって上下左右 4 方向に死角指向性を有する空間フィルタ群を形成する（図 -4）．

$$B_1(\omega)=X_1(\omega)-e^{-j\omega\tau}X_4(\omega)$$
$$B_2(\omega)=X_2(\omega)-e^{-j\omega\tau}X_1(\omega)$$
$$B_3(\omega)=X_3(\omega)-e^{-j\omega\tau}X_2(\omega)$$
$$B_4(\omega)=X_4(\omega)-e^{-j\omega\tau}X_3(\omega)$$

上下左右 4 つの空間フィルタの出力の振幅成分のうち，最も小さな成分を選択することで指向性雑音の成分を最も小さくした出力を得る．

$$|M(\omega)|=\min_i[|B_i(\omega)|]\quad(i=1,2,3,4)$$

最小値選択された成分から，さらに正面，すなわち目的音方向に死角指向性を持つ空間フィルタ成分 $N(\omega)$ を周波数減算することにより，目的音方向だけを残した成分 $Y(\omega)$ を得る．

### ●拡散性雑音抑圧

拡散性雑音抑圧は指向性雑音の抑圧と同じ 4 つの空間フィルタ出力を用いたマルチチャネルウィーナーフィルタ [4] で実現する．目的音である話者の声は各マイクロホンで観測される信号の相関が高いが，拡散性の雑音は観測信号間で相関が低い．この性質を利用し，対向する方向に指向性を持った信号を組み合わせ，互いの相関の程度を反映した係数を持つフィルタ $H_m(\omega)$ を構成する．

$$H_m(\omega)=\frac{|B_1(\omega)B_3^*(\omega)|+|B_2(\omega)B_4^*(\omega)|}{\frac{1}{2}\sum_{i=1}^{4}|B_i(\omega)|^2}$$

上式は分子のクロススペクトルを分母のパワースペクトルで正規化する形になっており，相関が高ければ 1 に，低ければ 0 に近づく特性を持つ．このフィルタを前記の指向性雑音を抑圧した信号 $Y(\omega)$ に乗じることにより，相関が低い成分を抑圧し拡散性雑音を低減する．

$$|\hat{S}_m(\omega)|=H_m(\omega)|Y(\omega)|$$

### ●残留雑音抑圧

指向性雑音，および拡散性雑音を抑圧した信号 $|\hat{S}_m(\omega)|$ に対し，さらにシングルチャネルのウィーナーフィルタ $H_s(\omega)$ を適用して残留する定常雑音を抑圧する．ウィーナーフィルタを適用するためには，非音声区間を検出して残留雑音のパワーを推定する必要がある．ここでは非発話区間検出に前段のマルチチャネルウィーナーフィルタの値が利用可能であるため別途発話区間推定を行う必要がないことも本方式の特徴である．

$$|\hat{S}(\omega)|=H_s(\omega)|\hat{S}_m(\omega)|$$

最終的に得られた振幅スペクトルに入力信号の位相を付与し，逆フーリエ変換することで雑音が抑圧された音声信号を復元する．

## 音源分離モジュールの開発

開発した音源分離方式を実環境で利用・評価するため，音声分離小型モジュールを開発した [2]

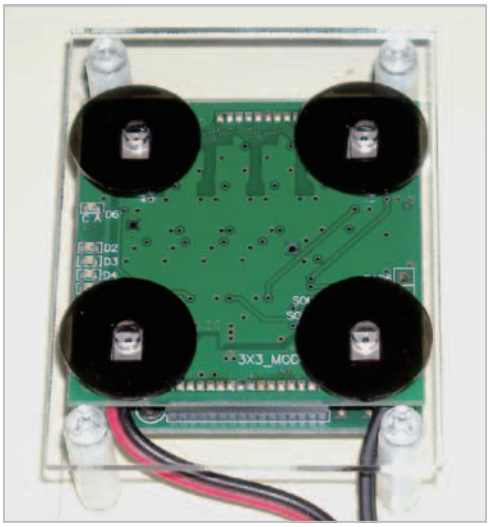

図 -5　音源分離モジュール

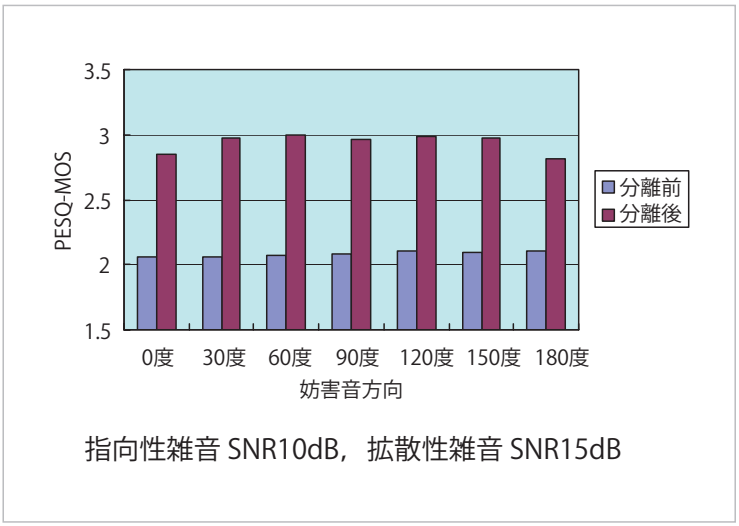


図 -6　分離性能評価

(図 -5)．モジュールは FPGA によって構成され，4 チャネルの MEMS マイクロホン，AD 変換器を搭載している．マイクロホン間の距離は縦横ともに 3cm と非常に小型であり，リモコンや携帯電話などの小型の機器にも実装可能である．内部では，4 個のマイクロホンの入力信号を標本化周波数 64kHz でオーバーサンプリングした後，16kHz にダウンサンプルする．その後，1024 サンプル(64ms)を分析単位(フレーム)として FFT の他一連の音声分離処理を行う．フレーム更新周期は 16ms であり，フレーム長の 64ms と併せて処理遅延は 80ms となる．分離音は DA 変換器を通してアナログ信号として出力され，音声分離モジュールが，いわば雑音抑圧機能を持ったマイクロホンとして機能する．そのため，従来のマイクロホンを使っていた音声認識装置などの機器に，そのまま接続できる構成となっている．

## 音源分離実験

方式評価用に，実際に試作機に実装された正方形マイクロホンアレイを用いて目的音，指向性雑音，拡散性雑音の収録・収集を行った．目的音は，携帯端末を手に持ち音声を入力するシーンを想定して，試作機正面 30cm の位置からスピーカー出力した．指向性雑音は，床から試作機と同じ高さで試作機に対して 1m の距離から，正面を 0 度として左回りに 0 度から 180 度まで，30 度ごとにスピーカー出力した．拡散性雑音としては，展示会騒音，道路騒音，車内騒音(高速道路走行，一般道路走行)などを実環境にて収録した．それぞれの収録音を目的に応じて所定の SNR のもとに混合し，方式の実証・評価に使用した．

混合音に対する分離性能を PESQ[5)，☆1] を使って評価した．指向性雑音 SNR10dB，拡散性雑音 SNR15dB における評価結果を図 -6 に示す．目的音と妨害音が同一線上に並ぶ場合を除いて，混合音に対して 0.8 以上向上し，ほぼ PESQ-MOS 値 3.0 を達成している．

## まとめ

上下左右に死角指向性を形成する 4 種の空間フィルタとマルチチャネルウィーナーフィルタ，シングルチャネルウィーナーフィルタの組合せにより，指向性雑音，拡散性雑音を同時に抑圧する，正方形マ

☆1　国際電気通信連合 ITU-T P. 862 で規定された客観的音質評価尺度．原音声と符号化などの処理により劣化した信号を比較し，5 段階主観評価(5：非常に良い，4：良い，3：まあ良い，2：悪い，1：非常に悪い)の MOS 相当値を推定する．



イクロホンアレイによる音源分離技術を開発した．

実環境騒音，および実機による入力データによる評価・方式改良を行い，SNR10dB の指向性雑音，および SNR15dB の拡散性雑音の重畳環境において，約 80ms の遅延で，分離音に対して PESQ-MOS 3.0 の品質を与える音声分離システムを，マイク間隔 3cm×3cm という非常にコンパクトな配置で実現した．

一方，実機の試作・評価を通して，いくつかの課題も見えてきた．空間フィルタは 1 対 2 個のマイクロホンからの入力を利用するが，一般にマイクロホンは製造誤差などにより感度が異なることがある．実機においてシミュレーションと同等の性能を実現するには，特性のバラツキを個体に応じて自動的に補正・正規化する処理が必要である．

提案方式では，目的音は正面から到来するものと仮定し，正面方向に鋭い指向性を形成することで音源分離を実現している．そのため正面から外れた目的音に対して抑圧や変形といった問題が生じる．利用形態によっては目的音方向に対するロバストネス向上のため，音源方向追尾の仕組みが必要となる．

また，提案方式では，各マイクロホンに入力される信号の位相情報が重要な役割を果たしている．したがって，筐体内部の音の反射や回りこみが生じないように，マイクロホンの実装上の注意が求められる．

以上のような課題に対しては，すでに実機上に対策を組み込み，雑音抑圧に関しては性能確保の目処が立っている．反面，目的音の歪やミュージカルノイズ☆2 の発生などの聴感上の音質には，やや課題が残されている．今後，一層の音質改善を行うことにより音声認識の前処理としてだけではなく携帯端末，会議端末などの通信機器への応用展開を図る．

☆2　周波数領域の非線形処理に伴う人工的な雑音．雑音成分の引き残し，その他の原因によって特定の周波数で信号が現れたり消えたりする現象が起き，処理後の音にキュルキュル，ピロピロといった耳障りな雑音が混入する．

**参考文献**


1) Ogawa, T., Takada, S., Akagiri, K. and Kobayashi, T. : Speech Enhancement Using a Square Microphone Array in the Presence of Directional and Diffuse Noise, IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences, Vol.E93-A, No.5 (2010).
2) 古井貞熙，小林哲則，矢頭　隆，大淵康成，河村聡典，三木清一，庄境　誠：(総合報告) 音声認識実用化技術の展開，電子情報通信学会誌，Vol.93, No.8, pp.725-740 (Aug. 2010).
3) Takada, S., Kanba, S., Ogawa, T., Akagiri, K. and Kobayashi, T. : Sound Source Separation using Null-Beamforming and Spectral Subtraction for Mobile Devices, 2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2007), MP1-04 (Oct. 2007).
4) Zelinski, R. : A Microphone Array with Adaptive Post-filtering for Noise Reduction in Reverberant Rooms, Proc. ICASSP, Vol.5, pp.2578-2581 (1988).
5) ITU-T Recommendation P. 862, Perceptual Evaluation of Speech Quality(PESQ) : An Objective Method for End-toEnd Speech Quality Assesment of Narrow-Band Telephone Networks and Speech Codecs (2001).


(平成 22 年 9 月 15 日受付)


**矢頭　隆**　yazu752@oki.com
1979 年九州大学工学部電気工学科卒業，同年沖ソフトウェア（株）入社．1980 年沖電気工業（株）入社．音声符号化，音声合成，音声認識の研究，および音声合成の製品開発などに従事．日本音響学会会員．

**森戸　誠**
1974 年東京工業大学工学部電子工学科卒業．1976 年同大総合理工学研究科物理情報専攻修了．同年，沖電気工業（株）入社．音声，音響技術の研究開発に従事．2009 年同社退職．

**山田　圭**　yamada648@oki.com
1998 年九州大学大学院システム情報科学研究科修了，2000 年同大学院博士課程後期中退．2001 年沖電気工業（株）入社．暗号回路，音源分離の研究開発，およびオーディオコーデックの製品開発などに従事．

**小川 哲司**（正会員）ogawa@pcl.cs.waseda.ac.jp
2005 年早稲田大学大学院理工学研究科博士課程修了．早稲田大学理工学部助手，同大学客員講師を経て，2007 年より早稲田大学高等研究所助教．博士（工学）．音声認識，音響信号処理に関する研究に従事．電子情報通信学会，日本音響学会各会員．