

音声認識や携帯電話などへの高品質な音声入力を目的として、ステレオのオーディオ入力を利用できる 2 チャンネル (ch) 適応マイクロホンアレーを開発した。このマイクロホンアレーは、アンテナなどの指向性制御に用いられる適応ビームフォーマを二つ並列動作させて話者方向を追尾し、適応処理で問題となっていた目的信号除去による性能低下を防止する。追尾に必要な話者方向は、目的信号を抑圧するように制約をおいたビームフォーマから検出しており、従来困難だった 2ch による追尾処理を可能にした。このマイクロホンアレーにより、移動する音源の信号を高精度で入力でき、話者位置の制約が緩和されて使い勝手の向上が期待できる。

This paper describes a two-channel adaptive microphone array which can utilize a stereo audio input. This microphone array tracks a target speaker by using two adaptive beam-formers, and can reduce target signal cancellation in cases where the target direction differs from the look direction of the beam-former. Target tracking using a two-channel microphone is made possible by estimating the direction of the target from the directional response of the beam-former, which is constrained to have a unity gain in the direction of arrival (DOA) of noise.

The proposed microphone array can extract signals from a moving target with high accuracy.

### 1 まえがき

ハンズフリーの音声認識や携帯機器への音声入力を目的として、複数のマイクロホンで受信した音を処理して雑音を抑圧するマイクロホンアレー処理が検討されている。なかでも、自動的に雑音方向に感度の低い死角を形成して雑音を抑圧する適応ビームフォーマを利用したものは、少ないマイクロホン数で高い性能が得られるため、パソコンや携帯機器など、ハードウェア上の制約から音声入力のチャンネル数が限られる場合に有利である。しかし、処理の設定と実際の使用環境との間のずれによって性能が劣化しやすく、マイクロホン数が少ない場合にこれを回避するのが困難なため、広範囲に応用されるには至っていない。

一般的な適応ビームフォーマは、処理に先立って目的とする信号の到来方向を設定し、処理中は目的の方向から到来する信号を通過させ、他の音源方向には死角を形成して信号を減衰させる。このとき、設定した方向と実際の目的信号の到来方向とがずれていると、目的信号が雑音とみなされて除去され性能が大幅に劣化する。性能劣化要因のなかでも目的信号の除去による影響がもっとも大きく、これを抑えるため、種々のロバスト化手法が開発されている(1),(2),(3)。

従来のロバスト化手法は、主に、到来方向のずれを許容するようにフィルタの指向性に拘束を設けるものであり、この拘束によりフィルタの自由度が減少して性能が低下するため、もともと自由度の小さい 2 ch アレーには適用困難

である。

これに対し、信号の到来方向を追尾する方法<sup>(2)</sup>が提案されており、フィルタの自由度を損なわずに到来方向のずれを減少させることができる点で有利である。しかし、追尾開始時に音源方向の初期値を与える必要があり、妨害雑音源のある場合にこれを 2 ch の入力から推定するのは困難である。また、非定常な信号である音声は信号が断続するため、信号途絶時に音源方向が変化した場合の対策も必要になる。

そこで、2 ch の場合でも到来方向を推定してビームフォーマの目的方向を話者に追従させ、実質的に到来方向の許容範囲を大きくしてロバスト性を高める方法を開発したのでこれについて述べる。

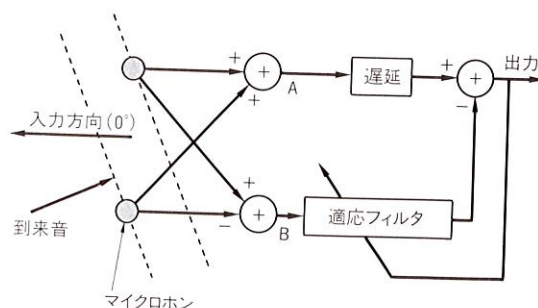


図1. 2ch Griffith-Jim GSC 目的音の到来方向とビームフォーマの入力方向が異なる場合、差分演算により目的音が完全に除去できないため、適応フィルタにより目的音が除去されて性能が低下する。

Two-channel Griffith-Jim GSC

## 2 2ch 適応ビームフォーマ

適応ビームフォーマは種々の方式が提案されているが、ここでは、演算量の点で有利な Griffith-Jim GSC (一般化サイドローブキャンセラ)<sup>(4)</sup>を用いて検討した。図1に2chの場合の Griffith-Jim GSCの構成を示す。なお、以下では、混乱を避けるため、ビームフォーマの目的方向を入力方向と呼ぶことにする。

図1のように、2chのGriffith-Jim GSCでは、二つの入力信号の差の信号(図中のB)を参照信号とし、入力信号を加算した信号(図中のA)を希望応答として適応フィルタによる雑音除去を行う。図中の入力方向から音が到来する場合は、音が二つのマイクロホンに同時に到達するので、2ch間の信号の差をとることによって目的音の成分が除去され、雑音成分だけが残る。また、2chの信号の和をとることにより、目的音の成分が強調される。そこで、差分により残った雑音成分を参照信号として適応フィルタにより雑音除去を行い、雑音と相関のある成分を和の信号から取り除いている。

一方、目的音の到来方向が入力方向と一致しない場合、入力信号間の差をとっても目的音成分が残るため、適応フィルタによって目的音が除去され出力信号が劣化する。残留成分がわずかであっても、適応フィルタにより大部分の目的音が除去されるので大きな性能劣化となる。マイクロホンアレーを音声収録に用いる際は、話者の動きにより音源の方向が一定しないことや、音源が安定した点音源でないことなどから、目的信号の除去が生じやすい。

音源が安定した点音源でないことに対しては、目的音の到来を検出して目的音入力中は適応を停止することにより対処できる。目的音検出方法として、マイクロホン信号間の相関と適応フィルタの出力パワーとに基づく適応速度の制御方法<sup>(3)</sup>が提案されているが、話者位置の移動の影響まで吸収することはできない。

## 3 二つのビームフォーマを用いた話者方向追尾

話者に合わせてビームフォーマの目的方向を変えるには、入力音声から話者方向を推定する必要がある。2ch入力の場合、従来の方向推定法は音源が1個の場合にしか適用できないが、2chマイクロホンアレーは、目的音のほかに1個の雑音源がある場合にこれを除去できることが特長であるため、音源2個の場合に方向推定する必要がある。そこで、一般的な方向推定法ではなく、ビームフォーマのフィルタの特性を利用して音源方向を推定する。

ビームフォーマのフィルタは、雑音方向に感度が低くなるように自動的に指向性を形成するため、指向性のうえて極小値を示す方向を雑音方向とすることができる。通常、

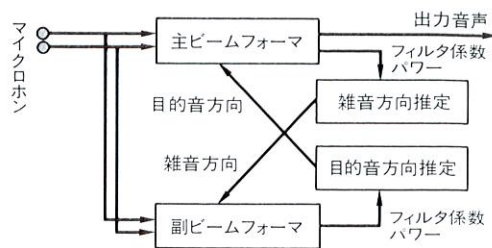


図2. 話者追従型2ch 適応マイクロホンアレー 二つの適応ビームフォーマにより目的音源と雑音源の方向を推定して追尾する。

Two-channel adaptive microphone array with target tracking

ビームフォーマは入力方向に対して感度を一定に保つように制約をおいているので目的音の方向には死角ができないが、雑音方向に入力方向を設定した別のビームフォーマ処理を行えば、目的音方向を求めることが可能である。二つのビームフォーマを使った2ch雑音抑圧処理の構成を図2に示す。

図2に示すように、マイクロホンから入力した信号を二つの適応ビームフォーマで並列に処理する。ビームフォーマの一方は雑音を抑圧して目的信号を出力する主ビームフォーマ、もう一方は雑音方向に入力方向を設定し、目的音を抑圧する副ビームフォーマである。

この構成において、主ビームフォーマのフィルタから雑音源の方向を推定して副ビームフォーマの入力方向とし、副ビームフォーマのフィルタから目的音の方向を推定して主ビームフォーマの入力方向とする。入力方向の設定は入力信号に遅延を与え、0°方向から信号が到来したのと等価になるようにすることにより行う。方向推定と入力方向設定を短い時間間隔で繰り返すことにより話者方向を追尾する。

なお、音源方向の推定は、雑音の影響を軽減するため、現在の瞬時の推定値にビームフォーマの出力パワーによる重みをかけて過去の推定値との間で平滑化し、現在時刻の音源方向とするようにしている。

## 4 性能評価

提案したマイクロホンアレーの性能評価のため、①音源方向対感度の特性、②移動音源の追尾性能、③移動音源信号の抽出性能について計算機シミュレーションにより検討した。

### 4.1 音源方向対感度の特性

適応マイクロホンアレーは、雑音源の方向に依存して指向性を変えるため、一つの場面設定に対する指向性だけからは特性を把握しにくい。さまざまな雑音源方向に対応した特性を調べるため、1個の音源の方向を固定して適応処理収束後の出力パワーを求める計算を、この音源の方向を順次変えながら行った。

ビームフォーマは、先に述べた 2 ch 間の相関と出力パワーに基づく適応速度制御を併用しており、適応速度制御と追尾処理を行わない場合、および適応速度制御と追尾処理を併用して追尾範囲を変えた場合とについて調べた。マイクロホン間隔は 16 cm とし、音源の SN 比 (信号と雑音の比) は周囲雑音に対し 25 dB とした。結果を図 3 に示す。

図 3 の曲線は、入力方向 0° に音源があるときの出力パワーで正規化してあり、(a) は音源が白色雑音の場合、(b) は音源が音声の場合である。図において、曲線 A は適応速度制御と追尾処理なしの場合、曲線 B, C, D は適応速度制御を行い、かつ、追尾範囲  $\phi$  を入力方向の周りに 0°,  $\pm 20^\circ$ ,  $\pm 40^\circ$  とした場合である。

図からわかるように、適応速度制御なし(A)の場合は目的信号の到来方向のずれに対する許容範囲が非常に狭いのにに対し、適応速度制御を行って音源追尾しない(B)の場合はずれの許容範囲が  $\pm 8^\circ$  程度に広がる。追尾する(C), (D)の場合は追尾範囲内にある音源の信号を減衰せずに入力できるようになるため、設定した追尾範囲にわたって感度が一定となる。これにより、許容範囲を任意の幅に設定できるようになることがわかる。

また、(a)と(b)を比較すると、音源が音声の場合のほうが若干許容範囲以外に対する感度を低く抑えることができる

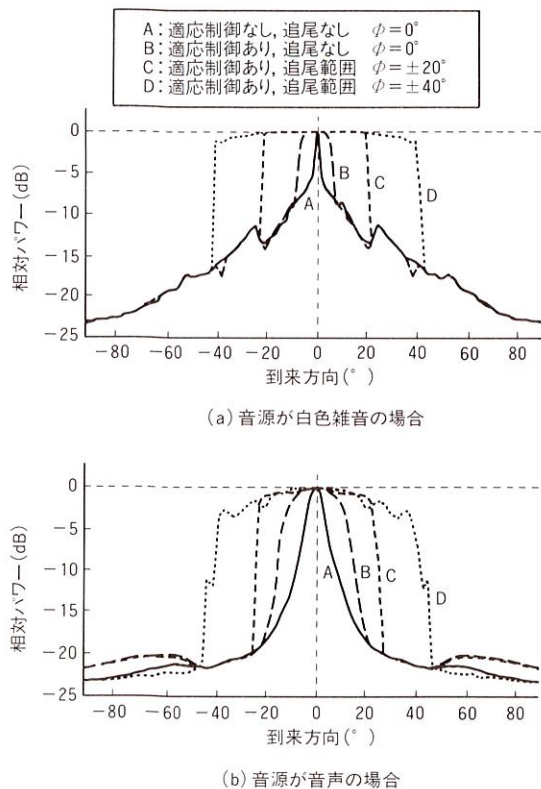


図 3. マイクロホン アレーの音源方向対感度の特性 設定した追尾範囲にわたって感度が平坦となり、ロバスト性が増す。

Sensitivity of microphone array vs. DOA

が、白色雑音の場合のほうが、追尾範囲の境界が急峻(しゅん)であり、境界に近い雑音源も抑圧できることを示している。

#### 4.2 移動音源の追尾

4.1 節は固定した音源に対する特性を調べた結果であるが、移動する音源に対する追尾性能を調べるため、音源を移動させて追尾軌跡を求めた。マイクロホンと音源は図 4 に示す配置とし、目的音源の方向を入力方向 0° の周りに正弦波的に移動するようにした。

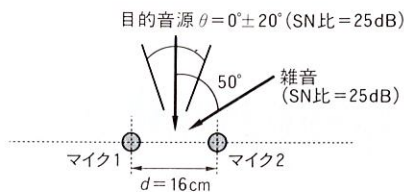


図 4. マイクロホンと音源の配置 目的音源の方向を  $\pm 20^\circ$  の範囲で正弦波的に移動させた。音源の SN 比は、背景雑音に対する値である。 Microphones and DOA locations

目的音源は単語発声、雑音源は文章読上げの音声とし、音源の SN 比は、周囲雑音を基準として 25 dB とした。この場合、目的音源対雑音のトータルの SN 比は  $-1.4\text{ dB}$  となった。目的音の方向は、3 単語の発声区間をセグメントとしてその間を固定し、セグメントごとに次々に従って方向を変化させた。

$$\theta = D \sin(k \times v),$$

ここで  $k$  はセグメントの番号、 $v$  は方向移動の角速度、 $D$  は方向変化の振幅であり、 $v = 30^\circ$ ,  $D = 20^\circ$  とした。

音源方向追尾結果を図 5 (a) に、目的音源と雑音源の波形を図 5 (b), (c) に示す。図 5 (a) において、実線は推定された目的音源の軌跡、点線は真の軌跡である。

図 5 (a) からわかるように、追尾誤差は一部を除いて数度以内であり、4.1 節で見たように適応速度制御によって許容される範囲が  $\pm 8^\circ$  程度であることを考慮すると、目的音除去は起こらず、十分な追尾性能であることがわかる。

#### 4.3 移動音源の信号抽出

最後に、目的音源が先と同じ条件で移動する場合について、目的音声の抽出性能を調べた。この結果、抽出した目的音の SN 比は、音源を追尾しない場合 4.2 dB、音源を追尾する場合には 14.7 dB となった。追尾しない場合は目的音が除去されるため大幅に性能が低下しているが、追尾した場合は問題なく目的音を抽出できており、従来法では達成できない移動音源の信号抽出が高精度で行えることを確認できた。

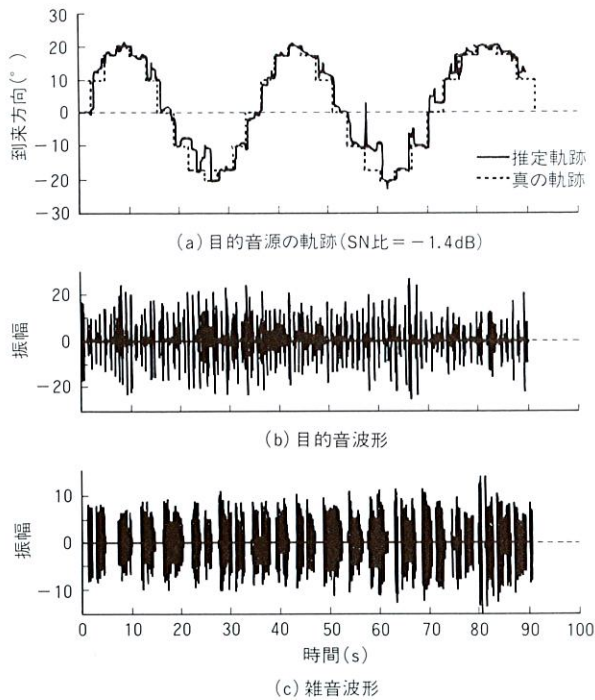


図5. 目的音源追尾結果 追尾軌跡の誤差は、適応停止により許容される範囲以内であるため、目的信号除去の現象は抑えられている。

Results of target tracking

## 5 あとがき

パソコンや携帯機器など、マイクロホン数が2に限定された場合でもロバストな処理が行える適応マイクロホンア

レーを開発した。計算機シミュレーションによる検討の結果、二つのビームフォーマを用いた音源追尾処理により、広い角度範囲からの到来音を高いSNで入力できることを確認し、従来雑音抑圧効果がほとんどなかった移動音源からの信号の抽出性能を大幅に改善できた。従来の適応マイクロホンアレーでは、使用の際、話者位置を固定しなければならなかったが、このマイクロホンアレーでは固定しなくて済むため、使い勝手の向上が期待できる。今後、さらに高雑音の環境にも対応できるように、多チャンネル化を検討していく。

## 文献

- (1) 宝珠山治, 他: ブロッキング行列にリーク適応フィルタを用いたロバスト一般化サイドロープキャンセラ, 電子情報通信学会論文誌 A, **J79-A**, 9, pp.1516-1524 (1996)
- (2) S. Affes, et al: Robust Adaptive Beamforming via LMS-like Target Tracking, Proc. Int. Conf. Acoust. Speech & Signal Process, pp.IV-269-272 (1994)
- (3) J.E. Greenberg and P.M. Zurek: Evaluation of an Adaptive Beamforming Method for Hearing Aids, J.Acoust. Soc. Am., **91**, 3, pp.1662-1676 (1992)
- (4) L.J. Griffith and C.W. Jim: An Alternative Approach to Linearly Constrained Adaptive Beamforming, IEEE Transactions on Antennas Propagat., pp.27-34 (1982)



永田 仁史 Yoshifumi Nagata, D.Eng.

研究開発センター 関西研究所主務, 工博。  
音声認識, 雑音抑圧処理の研究・開発に従事。電子情報通信学会, 日本音響学会, 情報処理学会会員。  
Kansai Research Lab.