

# クラスタ計算機を用いた平面アレーシステムによる 複数音源の分離抽出\*

– ANF 処理におけるフィルタ特性に関する検討 –

木下 智恵\*<sup>1</sup> 永座 強\*<sup>2</sup> 苅木 禎史\*<sup>1</sup> 宇佐川 毅\*<sup>1</sup>

\*<sup>1</sup>(熊本大学工学部) \*<sup>2</sup>(熊本大学大学院自然科学研究科)

## 1. はじめに

人に優しい情報入力装置の一つとして音声認識システムが挙げられる。パーソナルコンピュータなどにおいて音声認識を行う際、ヘッドセットなどの使用による不便さを解消するために、マイクロホンアレーを用いた信号分離システムによる集音手法の構築が必要となっている。

これまでに音声認識装置のための集音方法として、ディスプレイに組み込まれた平面マイクロホンアレーによる音声抽出システムが提案されている [1]。そのシステムはまず、ブラインド信号分離手法の一つである AMUSE (Algorithm for Multiple Unknown Signals Extraction) 法 [2] により複数音源を大まかに分離する。また、DSA (Delay-and-Sum Array) 法により最大パワを持つ音源 (第 1 音源) の DOA (Direction of Arrival) を推定する。その第 1 音源の方向情報を基に AMUSE 法による分離信号におけるパーミュテーション処理を行う。さらに分離信号を利用して第 2 音源の DOA 推定を行う。次に推定された音源の DOA を基に反復エコー除去法に基づく逆フィルタリング手法 [3] を用いてアレーを構成することにより生成した Adaptive Notch Filter (ANF) [4] を利用し目的音源以外の信号成分の抑制を行なう。

この音声抽出システムの ANF 処理部分では、フィルタの周波数特性によって分離抽出音源の特定周波数領域が減衰するという問題があった。本報告では、音源位置によってマイクロホンの組み合わせを自動的に選択し、フィルタのディップ点 (減衰部分) を信号抽出周波数の帯域外に配置することにより特定周波数における特性劣化を制御する手法を提案する。

また、このシステムでは、コンピュータオペレータを主要な話者として音声を分離抽出することを目的としているため、実用化には

実時間で動作することが必要とされる。

本来、実時間処理システムを構築する際、小さく安価で高速な DSP を用いるのが一般的である。しかし、アルゴリズムの開発段階では適しているとは言えない。そこで、検証段階として、汎用性の高いパーソナルコンピュータを用いたクラスタシステムによる実時間信号処理システムの構築が提案されている [5]。本報告では、クラスタシステムを用いて処理を分散させることによって、平面アレーシステムにおける処理速度の高速化を検討する。

## 2. 平面マイクロホンアレーによる音源抽出システム

これまでに提案されてきたアレーシステムのアルゴリズムを図 1 に示す。本手法は大きく 2 つのステップから成り、受信信号をフレーム単位に分割して処理を行う。

### (STEP1)

処理 1 受信信号を 4 個の sub-array 毎に分割、同期化し AMUSE 法により Blind Deconvolution を行う。

処理 2 DSA (Delay-and-Sum Array) 法により、最大パワを持つ音源 (第 1 音源) の DOA を推定する。

処理 3 前フレームの角度情報を基に、現フレームの角度情報と比較し第 1 音源、第 2 音源を決定する。

処理 4 各 AMUSE による分離信号から、第 1 音源と 2 番目に大きなパワを持つ音源 (第 2 音源) に対応する強調信号のみを取り出し、Step2 への入力信号とする。

### (STEP2)

処理 1 第 2 音源に対応する AMUSE 法の分離信号を用い、DSA 法により第 2 音源の DOA を推定する。

\* Extraction of multiple sound sources using a two-dimensional microphone array system on cluster system

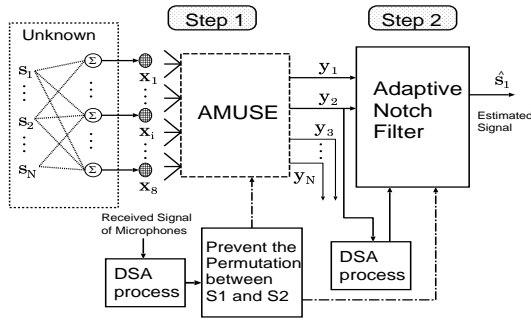


図 1. 従来のマイクロホンアレーシステムの構成  
 処理 2 第 1 音源に対応する AMUSE 法の出力信号と第 1, 第 2 音源信号をもとに, 構成した ANF を用いて目的信号の抽出を行う。  
 処理 3 ハニング窓を出力信号にかけて合成処理を行う。

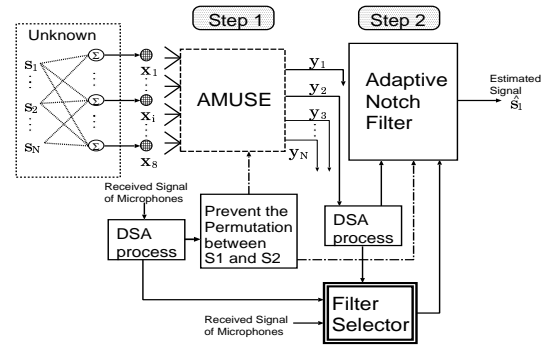


図 2. マイクロホン自動選択を取り入れたマイクロホンアレーシステムの構成

### 3. 抽出信号の特定周波数領域減衰に伴う ANF 処理に関する検討

従来提案された音声抽出システムの処理では, フィルタの影響により抽出信号の特定周波数域において減衰が生じるという問題があった [4].

そこで今回提案するマイクロホンの組み合わせを自動的に選択することにより特定周波数領域減衰を制御するアルゴリズムのブロック図を図 2 に示す。

この ANF 処理による複数音源の同時分離抽出では, 音源位置によりマイクロホン間の遅延が変化し, その影響によりフィルタの特性も変化する。よって, 各音源位置に対して目的周波数の成分が減衰しないようなマイクロホンの組み合わせを選択することでその問題を解決する手法を提案する。

ここでは, 2 音源をマイクロホン 2 本で抽出する場合を考える。このときの ANF 処理により抽出される信号の  $z$  変換  $\hat{S}_1(z)$  は次式で表される。

$$\hat{S}_1(z) = \frac{1}{2}(S_{sum}(z) - g(z)h^{-1}(z)\bar{s}_1(z))$$

$$S_{sum}(z) = 2S_1(z) + g(z)S_2(z)$$

$$g(z) = 1 + z^{(-\tau(\hat{\theta}_2, \hat{\phi}_2) + \tau(\hat{\theta}_1, \hat{\phi}_1))f_s}$$

$$\bar{s}_1(z) = h(z)S_2(z)$$

$$h(z) = 1 - z^{(-\tau(\hat{\theta}_2, \hat{\phi}_2) + \tau(\hat{\theta}_1, \hat{\phi}_1))f_s}$$

$f_s$  はサンプリング周波数,  $\tau(\hat{\theta}, \hat{\phi})$  は遅延差,  $S_{sum}(z)$  は全出力の同期加算,  $S_1$  は第 1 音源を  $z$  変換した信号,  $S_2$  は第 2 音源を  $z$  変換した信号, そして  $\bar{s}_1(z)$  は目的信号を含まない信号を示す。よって, 振幅特性  $|h(\omega)|$  は,

$$|h(\omega)| = \sqrt{2(1 - \cos(\omega\Delta\tau f_s))} \quad (1)$$

$$\Delta\tau = -\tau(\hat{\theta}_2, \hat{\phi}_2) + \tau(\hat{\theta}_1, \hat{\phi}_1)$$

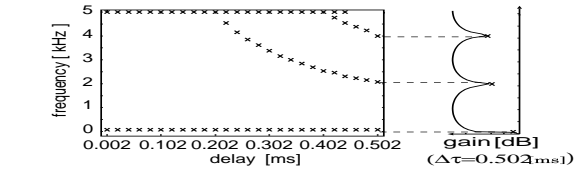


図 3. 各遅延におけるフィルタのディップ点と  $\Delta\tau=0.502[\text{ms}]$  におけるフィルタ特性

となる。式 (1) より, フィルタの特性は遅延成分  $\Delta\tau$ , サンプリング周波数  $f_s$  に依存し,  $\cos(\omega\Delta\tau f_s) = 1$  のときフィルタにディップ点ができることがわかる。この問題に対して特定の帯域が減衰しないような遅延差のマイクロホンの組み合わせを自動的に選択することによって周波数領域を制御する手法を提案する。

ここで,  $f_s=10\text{kHz}$  と設定し  $\Delta\tau$  を  $0.002\sim 0.502[\text{ms}]$  と変化させ, 生成されるフィルタの特性を求めた。各遅延におけるフィルタとディップ点の関係を表したものを図 3 に示し,  $\times$  はディップ点を表す。結果より, 目的とする周波数領域において減衰が生じることが確認できる。帯域  $100\text{Hz}\sim 3500\text{Hz}$  でディップ点を取らないようにするには遅延を  $\Delta\tau < 0.282[\text{ms}]$  に設定すれば良い。また, その条件を満たすマイクロホンの組み合わせは複数パターン存在するため, その中でも目的の帯域においてゲインの平均値が最大であるマイクロホンの組み合わせを自動的に選択する。

### 4. クラスタ計算機による処理の高速化

提案されている音声抽出システムの DOA 推定には, 従来の DSA 法を用いているが, この手法は演算量が多く, 処理に時間がかかるという問題点がある。そこで, この信号抽出システムの高速度処理について検討する。

今回, パーソナルコンピュータによって複数ノードを実現したクラスタシステムを用いた。クラスタシステムの構成を次に示す。

- マスターノード (1 台):  
Intel® Pentium® 4, 2.00GHz, 512MB

- スレーブノード :

Intel® Pentium®4, 1.60GHz, 512MB(7台)

Intel® Pentium®4, 2.00GHz, 512MB(12台)

ノード間でのデータの受け渡しはネットワークを介して行われるため、通信によるオーバーヘッドが生じる。また、動作環境は、OS:Red Hat Linux 7.3 2.96-110 上にて、SCore Ver.5.4.0 の通信用 MPI パッケージを用いた。

まず複数ノードへ分割する為の予備実験として、各ブロックに1ノードずつ割り当てて各処理時間を測定した。さらに、その各処理時間の割合から DOA 推定部分を分割しノードへの負荷を分散させた。

## 5. シミュレーション

まず、ANF 処理部分において特定周波数域が減衰する問題に対して、減衰しない遅延になるようなマイクロホンを選択する手法をシミュレーションにより検討する。次に、処理速度に関しては、クラスタ計算機を用いたときの処理速度について検討を行う。

シミュレーションには、14 inch 程度のディスプレイを想定して図4に示したような8個のマイクロホンを  $d_1 = 0.15\text{ m}$ ,  $d_2 = 0.10\text{ m}$  で並べた平面マイクロホンアレーを用いた。

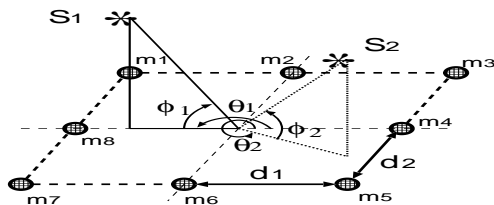
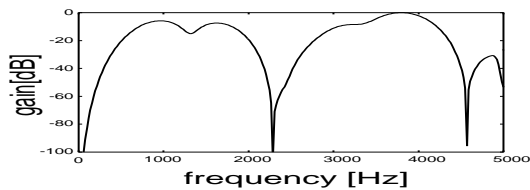
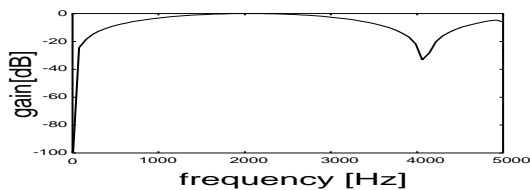


図4. マイクロホンの配置図



(a) マイク4つ使用した従来の ANF フィルタ特性



(b) マイク2つ使用した提案手法の ANF フィルタ特性

図5. ANF フィルタ  $h(z)$  特性

## 5.1 ANF 処理に関するシミュレーション

ANF 処理において、マイクロホンの組み合わせを自動的に選択することによって特定の周波数領域の減衰を制御するシミュレーションを行う。入力信号には、

第1音源 男声「大きな交差点があります」

$$\theta_1 = 240^\circ, \phi_1 = 60^\circ$$

第2音源 男声「会場はどちらですか」

音声に対して -5 dB のパワ、

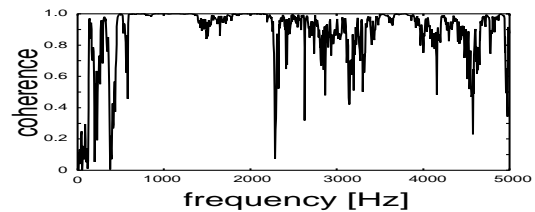
$$\theta_2 = 60^\circ, \phi_2 = 30^\circ$$

を用いる。ただし、角度を既知とする。ここで、第1音源、第2音源とディスプレイとの距離は  $L_1 = 0.5\text{ m}$ ,  $L_2 = 0.6\text{ m}$  とし、反復打ち切りレベル  $L_{th} = 30\text{ dB}$ , そしてサンプリング周波数は  $10\text{ kHz}$  とする。今回は電話の音声帯域  $100\text{ Hz} \sim 3500\text{ Hz}$  を確保できるように遅延を設定し、それに基づいてマイクロホンの組み合わせを自動的に選択する。

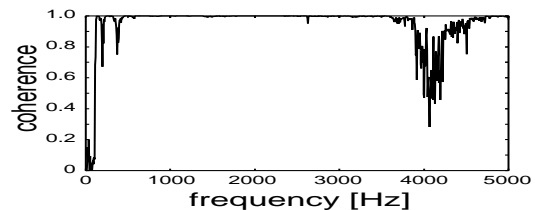
図5(a), (b) に従来法および提案手法の ANF フィルタ特性を示す。また、原信号と抽出信号とのコヒーレンスを図6(a), (b) に示す。

結果として、図5(a)に示す  $h(z)$ , 図6(a)に示す原信号と抽出音のコヒーレンスとの比較から、目的の周波数域が減衰する原因は ANF 処理において生成されたフィルタ  $h(z)$  によるものであることが確認できる。また図5(b)より、提案手法を用いることにより目的帯域においてディップ点がないフィルタが生成されたことがわかる。

図6(b)より、その目的とする周波数領域で品質が劣化しないようなフィルタを用いることで、目的信号も減衰せずに抽出できることが示された。



(a) 従来の ANF 処理によるコヒーレンス



(b) マイクロホンの自動選択を行った場合のコヒーレンス

図6. 原信号と抽出信号のコヒーレンス

表 1. ブロック毎の相対処理時間

第 1 音源の DOA 推定	136.9
AMUSE 法	3.49
第 2 音源の DOA 推定	44.8
ANF 処理	20.6

## 5.2 処理速度に関するシミュレーション

クラスタ計算機を用いることによってシステムの高速化を図る。

手法としてまず、システムを第 1 音源の DOA 推定部分, AMUSE 法部分, 第 2 音源の DOA 推定部分, ANF 処理部分に分け, 各 1 ノードにおける処理時間を測定する。次に, そのデータを基に DOA 推定部分の処理を分割し全体の処理速度について検討する。今回は, 図 1 のシステムにおいて処理を行う。

本報告では, フレーム長 2048tap, オーバラップは 1024tap, フレームシフトは  $t_{shift}=0.1024s$  に設定する。1 フレームの処理が  $t_{shift}$  内に終了するような処理を実時間処理と定義する。また, 音源は第 1 音源, 第 2 音源 (第 1 音源に対して -5dB のパワ) とともに, 男性話者による 2.0s の音声を用いた。

### 5.2.1 各ブロックの処理速度に関するシミュレーション

アレーシステムを各処理ごとに分割した場合の処理速度をシミュレーションにより検討する。表 1 に各ブロックに 1 ノードずつ割り当てた場合のブロック毎の相対処理時間を示す。ただし, 相対処理時間は, 実時間処理を 1 としたときの値であり, 全 18 フレーム中の最初と最後の 4 フレームを除いた 10 フレームの平均値から算出した。

表 1 のデータを基に第 1 音源の DOA 推定部分と第 2 音源の DOA 推定部分の分割の割合を決定する。処理を分割する際, マスターノードと AMUSE 法, ANF 処理は分割できないので各 1 ノードずつとなり, 残りのノードを DOA 推定部分で結果の割合により分割する。例えば, 20 ノードからなるクラスタ計算機を使用する場合は, 第 1 音源の DOA 推定に  $136.9 \times (20 - 3) / (136.9 + 44.8) = 13$  ノード, 第 2 音源の DOA に  $20 - 3 - 13 = 4$  ノードに割り振って処理を行うことになる。

### 5.2.2 処理ノード数と処理速度に関するシミュレーション

前シミュレーション結果を基にして, 平面アレーシステムをクラスタ計算機を用いて並列処理させたときの処理速度に関するシミュレーションを行う。

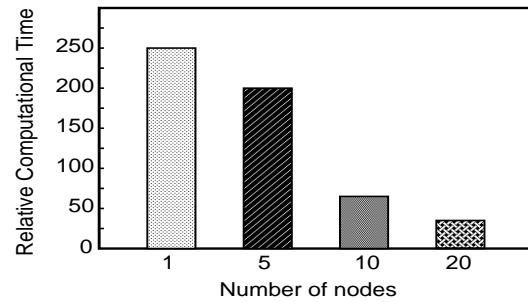


図 7. ノード数に対する相対計算時間

図 7 に処理ノード数に対するアレーシステムの処理時間の時間比を示す。ただし, 相対処理時間は, 実時間処理を 1 としたときの値である。結果よりノード数が増えるに従って処理速度も速くなっていることがわかる。また, ノード数 5 では各処理ごとに分割を行っているため処理時間が急速に短縮することはないが, ノード数 10, 20 の場合はさらに DOA 推定部分を分割しているため各ノードの処理時間が短くなり全体の処理速度が上がっている。しかし, まだ実時間処理を行う処理速度まで至っていない。

## 6. まとめ

本報告では, 以前から提案されている平面アレーシステムに関しての問題点を明確にし, 検討を行った。

抽出信号における特定周波数域の減衰について, シミュレーションによりその問題はフィルタの遅延成分などに依存していることを確認し, その遅延を適切な値に設定することで目的とする周波数帯域の減衰を解消できることを示した。

また, 処理速度に関するシミュレーションでは, クラスタ計算機を用いることにより処理の高速化を図った。シミュレーション結果からも処理速度が上がっていることが示されている。これにより, 効率よくシステムの構築ができると考えられる。しかし, 結果からもわかるようにまだ実時間で動くに至っていない。今後は, 実時間処理ができるようアルゴリズムの改良を行う。

## 参考文献

- [1] 宇佐川 毅他, 信学技報, EA2001-79(2001).
- [2] Lang Tong, *et al.*, Vol.38, No.5, pp.499-509, IEEE Trans. on CAS(1991).
- [3] 幸田 晃他, 日本音響学会誌, 53 巻 4 号, pp.268-276(1997).
- [4] Tsuyoshi Usagawa, *et al.*, Acoust. Sci. & Tech., Vol.22, No.4, pp.315-317(2001).
- [5] 荻木 禎史他, 日本音響学会講演論文集, pp.687-688(2002)