

NAIST-IS-DD0861015

博士論文

独立成分分析に基づく  
リアルタイムブラインド音源分離マイクロホン  
に関する研究

稗方 孝之

2009年11月5日

奈良先端科学技術大学院大学  
情報科学研究科 情報処理学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に  
博士(工学) 授与の要件として提出した博士論文である。

稗方 孝之

審査委員：

鹿野 清宏 教授 (主指導教員)

杉本 謙二 教授 (副指導教員)

猿渡 洋 准教授 (副指導教員)

# 独立成分分析に基づく リアルタイムブラインド音源分離マイクロホン に関する研究\*

稗方 孝之

## 内容梗概

近年，機械装置の音響診断，携帯電話，ロボット音声対話，カーナビ等，多くのアプリケーションで「聞きたい音」だけを瞬時に取り出したいニーズが高まっている．これら背景から，近年，音源分離技術の研究が盛んに行われているが，実環境における不十分な分離性能や演算量増加に伴うリアルタイム実装の困難さなどの問題が指摘されている．これを解決するため，本研究では，高速かつ高精度なブラインド音源分離技術の提案およびそのリアルタイム実装に関して論じる．

まずはじめに，雑音環境下で混合音を分離可能な新しいブラインド音源分離マイクロホンを提案・開発する．本マイクロホンは単一入力・複数出力型の独立成分分析とバイナリマスク手法を組合わせたリアルタイム処理向けのブラインド音源分離アルゴリズムを基礎としており，本アルゴリズムを改造してDSP（デジタルシグナルプロセッサ）上にリアルタイム実装する．さらに，本マイクロホンへのリアルタイム実装上の問題を詳説し，本マイクロホンの実験的評価によって，本マイクロホンの有効性を評価する．

次に，本マイクロホンをさらに現実的な環境で使えるようにするために，拡散性雑音環境下で特に有効な新しいブラインド音源抽出法を提案する．本手法は主マイクロホンと参照マイクロホンの各組の信号を基に並列に動作する複数の周波

---

\* 奈良先端科学技術大学院大学 情報科学研究科 情報処理学専攻 博士論文，NAIST-IS-DD0861015, 2009年11月5日.

数領域独立成分分析法，逆投影法，スペクトル減算を組合わせて構成されている．本論文では，これら手法の詳細について述べ，提案法を実装したマイクロホンを用いた様々な雑音環境下での評価実験から，提案法が従来法に比べて優れた性能を発揮することを示す．

キーワード

ブラインド音源分離, ブラインド音源抽出, DSP, リアルタイム実装, ポケットサイズマイクロホン

# A study on real-time blind source separation microphone based on independent component analysis\*

Takashi Hiekata

## Abstract

The real-time separation of target sound and noise is demanded in many applications, e.g., machine diagnosis systems, cellular phones, speech dialogue systems, and car navigation systems. Blind source separation (BSS) is beneficial for this purpose. Therefore, many BSS methods have been proposed. However, BSS implementation on a small-size LSI is very difficult because of computational complexities, or it is difficult to realize high separation performance in real-time on a small-size LSI.

In this paper, first, we develop a new blind source separation (BSS) microphone which can separate multiple sounds in noisy conditions. The BSS microphone is based on our previously proposed BSS algorithm, in which single-input multiple-output (SIMO)-model-based BSS and SIMO-model-based binary masking are combined. We modify this algorithm and implement it in a digital signal processor. Moreover, the issue of real-time implementation in BSS microphone is illustrated in detail, and the experimental evaluations of the hardware reveal the efficacy of the proposed BSS microphone.

Next, we proposed a new blind signal extraction (BSE) method in widespread

---

\* Doctoral Dissertation, Department of Information Processing, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD0861015, November 5, 2009.

i.e., more realistic noise conditions, which is based on multiple frequency-domain independent component analysis (FDICA) combining projection back (PB) and spectral subtraction (SS). We illustrate and evaluate the proposed method and BSE microphone. And experimental results reveal that the extraction performance of the proposed method is superior to that of conventional methods, using the BSE microphone in which proposed method is implemented.

**Keywords:**

blind source separation, blind source extraction, DSP, real-time implementation, pocket-size microphone

# 目次

1. 緒言	1
1.1 はじめに	1
1.2 従来研究とその問題点	3
1.3 本研究の狙い	4
1.4 関連研究と本研究の位置づけ	6
1.5 本論文の構成	7
2. 混合過程と従来の音源分離法およびリアルタイム実現例	8
2.1 まえがき	8
2.2 混合過程	8
2.3 従来の BSS	9
2.3.1 BM に基づく BSS	9
2.3.2 ICA に基づく BSS	10
2.4 従来のリアルタイム BSS システム	12
2.5 本章のまとめ	13
3. ポケットサイズブラインド音源分離マイクロホン	14
3.1 まえがき	14
3.2 SIMO-ICA と SIMO-BM を統合した 2 段型 BSS	14
3.2.1 アルゴリズム構成	14
3.2.2 周波数領域 SIMO-ICA (FD-SIMO-ICA)	16
3.2.3 SIMO モデルに基づくバイナリマスク	18
3.2.4 本構成の特徴	20
3.3 リアルタイム BSS マイクロホン	20
3.3.1 概要	20
3.3.2 リアルタイム実装	24
3.3.3 遅延時間	29
3.3.4 リソース	32

3.4	基本性能評価	32
3.4.1	評価値	32
3.4.2	実験条件	34
3.4.3	実験結果	34
3.5	リアルタイム処理による実験	36
3.5.1	移動音源による分離性能評価	36
3.5.2	分離性能のポラパターン	39
3.6	本章のまとめ	41
4.	並列ICAによる雑音推定に基づくブラインド音源抽出	42
4.1	まえがき	42
4.2	従来法	42
4.2.1	ICAに基づくBSS	42
4.2.2	ICAに基づくBSSとSSの単純接続	44
4.3	アプローチ	44
4.4	提案法	45
4.5	ブラインド音源抽出マイクロホン	49
4.5.1	概要	49
4.5.2	リアルタイム実装	52
4.5.3	リソース	54
4.6	音源抽出実験	55
4.6.1	実験条件	55
4.6.2	実験結果	58
4.7	本章のまとめ	68
5.	結言	69
5.1	まとめ	69
5.2	今後の課題	70
	謝辞	71



参考文献	73
付録	79
A. ICA による拡散性雑音推定の理論的考察と実験	79
A.1 ICA により形成されるフィルタの具体例 . . . . .	79
A.2 実験条件 . . . . .	81
A.3 実験結果 . . . . .	82
B. SIMO-ICA と SIMO-BM を結合した 2 段階 BSS 法を用いた音声認識 実験	84
B.1 実験条件 . . . . .	84
B.2 実験結果 . . . . .	85
発表リスト	86
特許リスト	93
プレスリリース	95

## 目 次

1	ハンズフリーシステム	2
2	アプリケーション例	2
3	関連研究と本研究の位置づけ	6
4	マイクロホンアレーの配置	9
5	周波数領域 ICA におけるブラインド音源分離の手順	11
6	$K = L = 2$ の場合の FD-SIMO-ICA と SIMO-BM を接続した提案 2 段型 BSS の信号の入出力の関係	15
7	BSS モジュール (1 号試作機)	21
8	BSS マイクロホン (SSM-001)	22
9	BSS マイクロホンのハードウェア構成	23
10	BSS マイクロホンの内部基板	23
11	リアルタイム処理の流れ	25
12	DSP と SDRAM の内部ブロック構成	26
13	入出力バッファの管理	30
14	遅延時間短縮のためのオーバーラップ加算 (8 倍の場合)	31
15	実験結果:(a) 雑音抑圧量, (b) ケプストラム歪, (c) PESQ MOS-LQO	35
16	実験で用いた残響室のレイアウト	37
17	リアルタイム BSS での実際の録音データを用いて計算した 0.5 秒毎のセグメンタル NRR	38
18	分離性能のポーラパターン計測条件	40
19	分離性能のポーラパターン計測結果	40
20	FDICA に基づく音源分離法	43
21	FDICA と SS の単純接続法	43
22	ICA による死角型フィルタ	46
23	提案法のブロック図	47
24	音源抽出マイクロホンの内部基板	50
25	音源抽出マイクロホンのハードウェアブロック図	51
26	提案法のリアルタイム処理フロー	53

27	目的音声と妨害音声の集音条件 . . . . .	56
28	拡散性雑音の集音条件 . . . . .	56
29	環境雑音を集音した駅環境 . . . . .	57
30	拡散性雑音を白色雑音 SNR 20 dB とした場合の音源抽出実験結果	59
31	拡散性雑音を白色雑音 SNR 10 dB とした場合の音源抽出実験結果	60
32	拡散性雑音を白色雑音 SNR 0 dB とした場合の音源抽出実験結果	61
33	拡散性雑音を音楽雑音 SNR 20 dB とした場合の音源抽出実験結果	62
34	拡散性雑音を音楽雑音 SNR 10 dB とした場合の音源抽出実験結果	63
35	拡散性雑音を音楽雑音 SNR 0 dB とした場合の音源抽出実験結果	64
36	拡散性雑音を駅雑音 SNR 20 dB とした場合の音源抽出実験結果 .	65
37	拡散性雑音を駅雑音 SNR 10 dB とした場合の音源抽出実験結果 .	66
38	拡散性雑音を駅雑音 SNR 0 dB とした場合の音源抽出実験結果 .	67
39	実験に用いた残響部屋のレイアウト . . . . .	81
40	分離実験のシミュレーション結果 . . . . .	82
41	空間的無相関雑音環境下における 1.5 kHz かつ 2 素子アレーの場合 の典型的な ICA フィルタの指向パターン . . . . .	83
42	話者位置と手法の違いによる単語認識精度の評価結果 . . . . .	85

## 表 目 次

1	従来の BSS 手法に関する利点・問題点 . . . . .	13
2	BSS マイクロホン諸元概要 . . . . .	24
3	提案法リアルタイム実装時のリソース . . . . .	32
4	音源 $2(\theta_2)$ の移動条件 . . . . .	36
5	音源抽出マイクロホン概略仕様 . . . . .	51
6	提案法リアルタイム実装時のリソース . . . . .	54
7	アプリケーション毎の課題 . . . . .	70
8	音声認識実験のパラメータ . . . . .	84

# 1. 緒言

## 1.1 はじめに

我々を取り巻く世界のあらゆる場所で音があふれている。街を例にとっても、人間の話し声、自動車や電車の走行音、機械の動作音など様々である。これら音は空気の振動として常に混ざりあった状態で人間の耳に到達する。

このような物理現象に対して、人間はどのように対処しているのだろうか。人間の持つ優れた聴覚能力の一つとして、音源分離というものがある。これは、それら混ざり合った音の中から「聞きたい音」だけを選び分けて聞くことができる能力のことを指す。特に、複数の話者が存在している環境で、特定の人物の声のみを聞き取れる能力は、カクテルパーティ効果と呼ばれている。その昔、聖徳太子は10人の話を同時に聞き分けたという逸話もある。

この「聞きたい音だけを聞く」という、人間の能力を電子機器や機械装置等で実現できないのであろうか。これが本研究に至る動機であり、本研究の最終目標である。本研究成果がもたらす社会貢献は非常に大きなものがあると予想される。例えば、一つの応用例は、図1のハンズフリー音声認識装置である。実現すれば、ユーザーはマイクロホンの位置を気にする必要がなく、手の拘束も受けないため、非常に使い勝手がよい。音声認識技術は、現状は接話型マイクロホンやヘッドセットマイクロホンを用いた近接入力を行わない限り、高精度な認識が行えない。本研究が当該分野のブレークスルーになる期待がある。他にも、図2に示す機械装置の音響診断、携帯電話での音声通話、ロボットとの音声対話、カーナビ等、様々なアプリケーションへの応用が期待できる。さらには補聴器への応用も考えられる。既に、この能力を模倣し、混ざりあった音の中から特定の音源信号を抽出する技術は音源分離技術あるいは音源抽出技術と呼ばれ、研究が盛んに行われているが、近年の計算機の進歩に従い、実現できる事は日進月歩で拡大している。

そこで、本研究では、聞きたい音だけを聞くことができる装置の実現を目標とする。

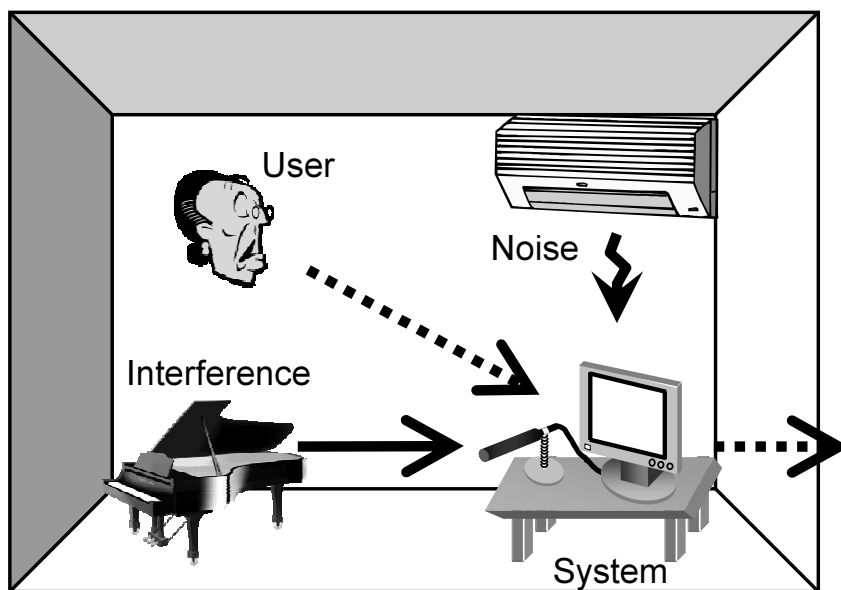


図 1 ハンズフリーシステム

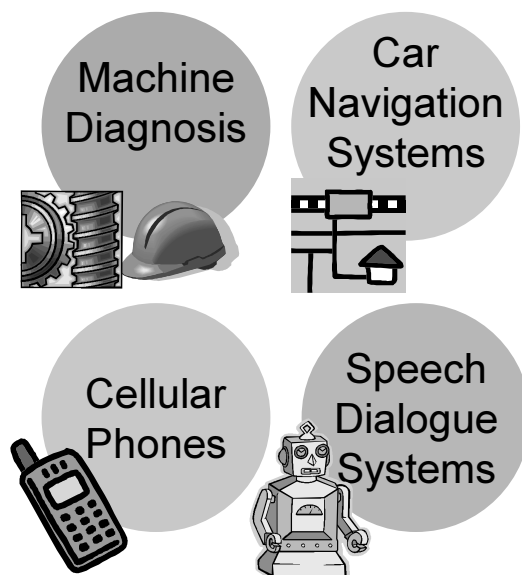


図 2 アプリケーション例

## 1.2 従来研究とその問題点

聞きたい音だけを聞く機能を実現する手段としては、既にいくつかの雑音抑圧手法が提案されている。単一マイクロホン入力に基づく代表的な例として、雑音の定常な振幅スペクトル成分の減算に基づくスペクトルサブトラクション (Spectral Subtraction; SS) などが提案されている [1]。SS は、非音声区間から雑音のパワースペクトルを推定し、雑音が加わった音声のパワースペクトルから、推定を行った雑音のスペクトルを減算することによって雑音の抑圧を行う手法である。しかし SS は、雑音の定常性を仮定しているために、非定常雑音には適用できないという問題点をもつ。また、こうした単一チャンネル信号処理は、雑音の抑圧に有益な音源の位置や到来方向といった空間的情報を用いて非目的音を効率的に抑圧することが困難であるという問題がある。

これに対して、多チャンネル信号処理としてマイクロホンアレーを用いたアレー信号処理 [2, 3] や、音源の独立性に基づく手法などが提案されている。アレー信号処理は、指向特性を形成することで音源分離を実現する技術であり、音源の空間的位置の違いによって非目的音を効果的に抑圧する技術である。遅延和アレー (DS) は、目的音の到来方位を用いて、その到来方位からの信号を同相化することで目的音を強調するアレー信号処理である。DS は、その構造が非常に簡単であるが、低周波数帯域での位相差情報を得るには、広い間隔のマイクロホンアレーが必要となる。その結果、高精度な音声強調を行うには大規模な装置が必要となる。

DS に比べ、小規模な構造で高精度な雑音抑圧が行える線形アレー信号処理として、適応型アレー (Adaptive Beamformer: ABF) が提案されている。ABF は、事前に目的信号が存在しない状況において非目的音を抑圧する指向特性を形成し、混合音から目的音のみを抽出するアレー信号処理である [3]。ABF は、比較的小規模なマイクロホンアレーにおいても高い分離性能が得られるという利点があるが、非目的音学習区間と目的音方位の 2 つの事前情報が必要となる。実際の応用においては、ABF はこの 2 つの事前情報が得られないことが多く、適用が困難であるという問題点がある。例えば、目的音方位が分からない混合系では、MUSIC 法 [4] などによって観測信号から目的音方位を推定することになるが、規模が小さく音源数よりマイクロホン数が大きくないアレーでは、目的音方位の推

定は困難である．また，雑音が大きい環境下では非目的音学習区間の検出も困難となる．これらの問題点は，ABF が教師あり学習に基づく信号処理であることに起因する．

これに対し，音源に関する事前情報が得られない状況下で，観測信号のみから音源分離を行う技術をブラインド音源分離 ( Blind Source Separation : BSS ) と呼ぶ．近年，音源信号が統計的に互いに独立であるという仮定のみを用いて，多チャンネルの観測信号のみから分離を行う独立成分分析 ( Independent Component Analysis : ICA ) に基づく BSS 手法が盛んに研究されている [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15] ．Cardoso らや Jutten らは高次相関の最小化による ICA を提案した [5, 6] ．Bell らは，相互情報量の最小化に基づく Infomax アルゴリズムを提案した [8] ．Amari らは，ICA の収束の安定性と速度向上のために自然勾配法に基づく ICA を提案している [9] ．また，Smaragdakis, Choi ら，Parra ら，Saruwatari ら，Nishikawa ら，Sawada ら，Araki らは，特に音響信号の BSS に特化した ICA を提案している [10, 11, 13, 15, 16, 17, 18, 19, 20, 21] ．これらの手法は，音源を事前情報なしに分離することを目的としているため，音源の種類によらず分離を行うことが出来るという利点を有している．特に，前記の従来研究では，比較的弱い残響環境下において数個 (2~3 個) 程度の点音源の分離に有効であるとの結論が示されている．一方，演算量の増加や統計量の推定精度の問題より，リアルタイム処理 (信号入力とほぼ同時に処理信号が出力できること) を導入することが困難であったり，複雑な音響環境への対応が不十分であるという欠点も従来より指摘されてきた．

### 1.3 本研究の狙い

本研究では，多チャンネル信号処理の枠組みにおいて，高い分離性能を持ちつつリアルタイムで動作する音源分離 ( BSS ) 装置および音源抽出 ( Blind Source Extraction : BSE ) 装置の実現を目標とする．その実現には，1) 分離・抽出性能に優れ，2) 雑音に頑健で，3) 瞬時に処理可能で，4) 手のひらサイズ，といった実用面を重視して，アルゴリズム面，ハードウェア面から実用的研究・開発を行う．

まずはじめに，BSS を行うアルゴリズムの提案およびその装置の開発を行う．BSS のアプローチは大きく 2 通りに分けられる．ICA に基づくアプローチとそれ以外の手法でのアプローチである．ICA に基づくアプローチ以外の手法としては，SS と似た手法であるバイナリマスク [31, 32, 33] が代表的である．これは処理が簡単なため，リアルタイム処理が求められるシステムに向いているという利点があるが，分離すべき音響信号同士の関係に制約があるなど様々な前提条件を必要とするため，高精度な分離処理は，一般的に困難であるといえる．一方，ICA に基づく BSS は，反復学習を用いて分離フィルタの非線形最適化を行っているために，高残響下においては局所解に陥り，完全な分離がなされずに分離信号中に非目的音が残留するという問題点があると共に，BSS は反復学習に時間がかかり，ハンズフリー音声認識システムに必要とされる，リアルタイム処理を行うことが困難であるという問題点がある．ICA のリアルタイム実装例は既にいくつか存在する [26] が，これらの全ては高速な計算機を必要とし，小型 LSI への実装例はほとんどない．また，アルゴリズムそのものがリアルタイム処理を意識していない．

これら背景から，本研究では，リアルタイム処理に適したブラインド音源分離手法として，多チャンネル出力に基づく ICA と，これらの ICA の分離信号に対してバイナリマスク処理を行う手法を採用し，高精度・高品質なリアルタイムブラインド音源分離装置の実現を目指す．

さらに，BSS 装置を発展させて，BSE を行うアルゴリズムの提案およびその装置の開発を行う．BSE 装置は，BSS 装置が機能するために必要な混合過程を必要としない点でより実用的といえる．つまり，BSS 装置は，目的音に加えて雑音も点音源を仮定していたが，BSE 装置ではこの混合過程を前提とせず，雑音が種々方向から到来する拡散性雑音という，より現実的な状況を仮定している．

これらの研究・開発を通じて実用的なリアルタイム BSS 装置，BSE 装置の実現を目指す．



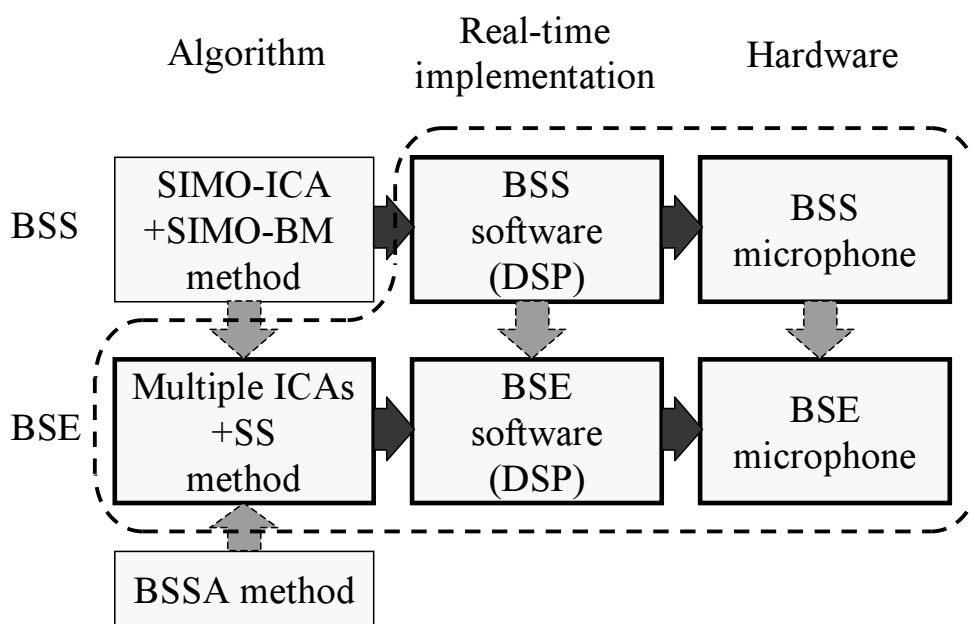


図 3 関連研究と本研究の位置づけ

#### 1.4 関連研究と本研究の位置づけ

本研究では、前述の通り、BSS 装置および BSE 装置の実現を目標とするが、その実現における関連研究との位置づけを図 3 に示す。まず、BSS 装置については、基盤技術となる基本アルゴリズムは、Mori ら我々の研究グループが考案した、3.2 章に記載の SIMO-ICA と SIMO-BM を統合した 2 段型 BSS [27] [28] をベースとし、BSS マイクロホンの研究に繋げる。また、BSE マイクロホンについては、Takahashi らの BSSA 法 [29] の研究の知見を生かしたリアルタイム実装向けの新たな手法を提案し、BSS マイクロホン研究・開発の知見と合わせて BSE マイクロホンの実現を目指す。よって、図 3 中の点線で囲われた範囲が本研究の対象範囲である。

## 1.5 本論文の構成

2 章では，実環境における混合過程と従来研究されている音源分離手法およびリアルタイム実現例について述べ，その問題点についてまとめる．次に 3 章では，採用する音源分離法，および開発した BSS マイクロホンについて説明を行い，その有効性を検証するための音質評価実験，リアルタイム処理実験について報告する．さらに，4 章では，拡散性雑音環境下というより現実的な環境で音源抽出が可能な新たな提案法および提案法を実装した BSE マイクロホンについて説明を行い，提案法および BSE マイクロホンの有効性を検証するための音源抽出実験について報告する．最後に，5 章にて本論文の内容をまとめ，本論文に記載した技術への思いと今後の課題を述べる．

## 2. 混合過程と従来の音源分離法およびリアルタイム実現例

### 2.1 まえがき

本章では，本研究において最初に前提として定義した混合系を示し，音源分離手法についての従来法についての説明を行う．まず 2.2 節では，本研究で問題として取り扱う混合系を定義し，その定式化を行う．次に，2.3 節では，従来の BSS に関して，BM に基づく手法と ICA に基づく手法に関して説明を行う．そして 2.4 節では，これら手法に関するリアルタイム BSS システムに関して説明を行う．

### 2.2 混合過程

本研究では，まずはじめに， $L$  個の音源から発せられた音源信号が空間を伝達して，直線上に配置された  $K$  個のマイクロホンで観測されるような系を考え（図 4 参照），音源数とマイクロ本数は等しい（ $L = K$ ）場合を考える．本研究では，特に（ $L = K = 2$ ）を取り扱う．

複数の混合信号はマイクロホンアレーにより観測され，観測信号は A/D コンバータを通して離散時間系列に変換される．離散時間フーリエ変換を適用することにより，観測信号は周波数領域において，複数の音源信号の線形混合と，ある加法性雑音で表現することができる．これを定式化すると以下のような式で表すことができる．

$$\begin{aligned}\mathbf{X}(f, t) &= \left[ X_1(f, t), \dots, X_K(f, t) \right]^T \\ &= \mathbf{A}(f)\mathbf{S}(f, t) + \mathbf{N}(f, t)\end{aligned}\quad (1)$$

$$\mathbf{S}(f, t) = \left[ S_1(f, t), \dots, S_K(f, t) \right]^T \quad (2)$$

$$\mathbf{A}(f) = \begin{bmatrix} A_{11}(f), & \dots, & A_{1L}(f) \\ \vdots & \ddots & \vdots \\ A_{K1}(f), & \dots, & A_{KL}(f) \end{bmatrix} \quad (3)$$

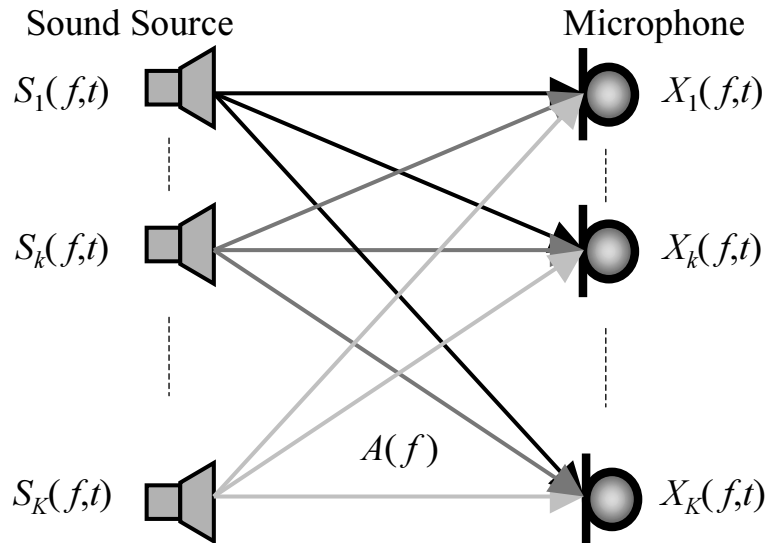


図 4 マイクロホンアレーの配置

$$\mathbf{N}(f, t) = \left[ N_1(f, t), \dots, N_K(f, t) \right]^T \quad (4)$$

ここで,  $S_l(f, t)$  ( $l = 1, \dots, L$ ) は音源  $l$  から発せられた音源信号,  $A_{kl}(f)$  ( $k = 1, \dots, K; l = 1, \dots, L$ ) は音源  $l$  からマイクロホン  $k$  までのインパルス応答を表し,  $X_k(f, t)$  ( $k = 1, \dots, K$ ) はマイクロホン  $k$  で收音される観測信号であるとする. また,  $N_k(f, t)$  ( $k = 1, \dots, K$ ) は加法性雑音を表す項目であり, 一般的には背景雑音やセンサ雑音等が含まれる. 混合行列  $A(f)$  は複素数であり, これは本研究で扱う混合過程のモデルが, 信号のマイクロホン間での時間遅れや部屋の残響時間を表現しなければならないためである.

## 2.3 従来の BSS

### 2.3.1 BM に基づく BSS

バイナリマスク [31, 32, 33] は教師なし BSS 手法の一つである. この手法は, 聴覚マスクング現象を模擬したものであり, より強い信号は弱い信号をマスクす

るという仮定に基づいている．バイナリマスクの決定は観測信号の振幅を比較することで推定を行い，耳（マイクロホン）に接近しているパワーの強い音源成分を選択的に抽出することで行う．この処理は時間-周波数領域で行われ，目的音源信号が支配的な領域はそのまま何も処理をせず，他の領域ではマスク処理を行う． $l$  番目の音源が  $l$  番目のマイクロホンに近いとの仮定の元に， $L = 2$  の時を考えると， $l$  番目の分離信号は次の式で与えられる．

$$\hat{Y}_l(f, t) = m_l(f, t) X_l(f, t) \quad (5)$$

ここで  $m_l(f, t)$  はバイナリマスク演算子で，以下の式で定義される．

$$m_l(f, t) = \begin{cases} 1 & |X_l(f, t)| > |X_k(f, t)| \quad (k \neq l) \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

この手法はごく少量の計算量しか必要としないため，リアルタイム処理が可能である．しかし，この手法は音源のスペクトル成分間にスパース性，つまり，時間-周波数領域において音源同士で成分の重なりが無いことを仮定している．ところが，一般的な音声信号の混合問題ではこの仮定は満たされることが多い（実際，音声と一般的な広帯域定常雑音は多くの成分において重なりを持っている）．そして，スパース性が成り立っていない状態では，この手法はひずみの一種であるミュージカルノイズと呼ばれる耳障りな雑音を発生する．

### 2.3.2 ICA に基づく BSS

本節では ICA に基づく BSS の詳細について述べる．

周波数領域 ICA (FDICA) [10, 12, 15, 16] では，まず初めに観測信号の短時間分析を，離散フーリエ変換 (DFT) を用いてフレーム毎に行う（図 5 参照）．各マイクロホン入力に対し，フレーム毎に周波数ビンにスペクトル値を記すと，時系列としてスペクトルを取り扱うことができる．これ以降，時系列を  $X(f, t) = [X_1(f, t), \dots, X_K(f, t)]^T$  と表すことにする．

次に，複素分離行列  $\mathbf{W}(f) = [W_{lk}(f)]_{lk}$  を用いて， $L$  個の出力時系列  $\mathbf{Y}(f, t) = [Y_1(f, t), \dots, Y_L(f, t)]^T$  が互いに独立になるよう，入力信号の分離を行う．ここ

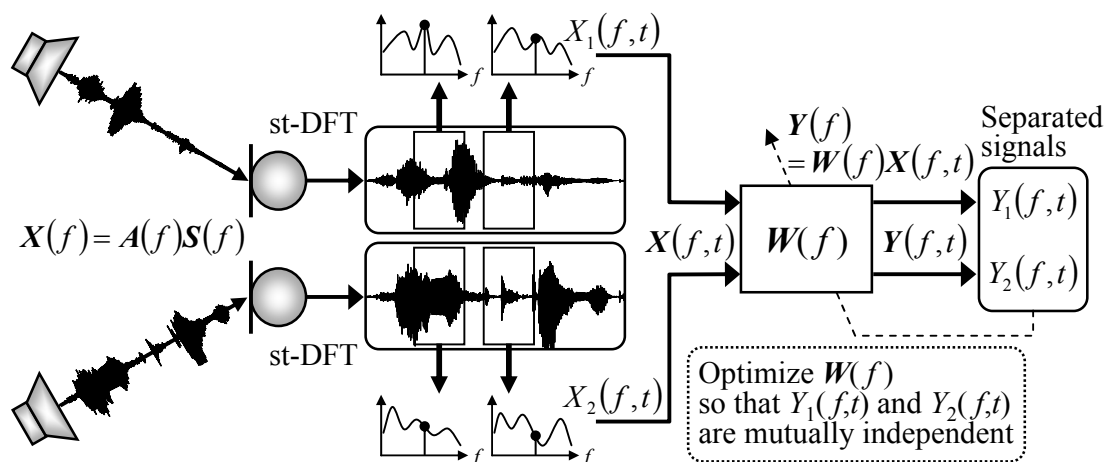


図 5 周波数領域 ICA におけるブラインド音源分離の手順

で,  $[\cdot]_{ij}$  は  $i$  行  $j$  列要素が  $\cdot$  である行列を表す. この処理は以下の式で定式化できる.

$$Y(f, t) = W(f) X(f, t) \quad (7)$$

$W(f)$  の最適化は ICA の手法により様々な種類が存在する. 例えば, Parra らにより提案されている, 2 次の統計量に基づく ICA (2nd-order ICA) では以下の更新学習式 [15] を用いる.

$$W^{[i+1]}(f) = -\eta \sum_{\tau} \alpha(f) \text{off-diag}(\mathbf{R}_{yy}(f, \tau)) \mathbf{W}_i(f) \mathbf{R}_{xx}(f, \tau) + W^{[i]}(f) \quad (8)$$

ここで,  $\eta$  は学習係数,  $\text{off-diag}[\mathbf{X}]$  は行列  $\mathbf{X}$  の全対角様子を 0 に置き換える演算子,  $[i]$  は  $i$  番目の反復における値,  $\alpha(f) = (\sum_{\tau} \|\mathbf{R}_{xx}(f, \tau)\|^2)^{-1}$  は正規化のための係数である ( $\|\cdot\|$  はフロベニウスノルムを表す).  $\mathbf{R}_{xx}(f, \tau)$  と  $\mathbf{R}_{yy}(f, \tau)$  はそれぞれ入力  $x(f, t)$  と出力  $y(f, t)$  のクロスパワースペクトルを表しており, 複数の時刻  $\tau$  の周辺で計算される.

一方, 高次統計量に基づく ICA (Higher-order ICA) では以下の更新式 [10] が

用いられる．

$$\begin{aligned} \mathbf{W}^{[i+1]}(f) = & \eta \left[ \mathbf{I} - \langle \Phi(\mathbf{Y}(f, t)) \mathbf{Y}^H(f, t) \rangle_t \right] \mathbf{W}^{[i]}(f) \\ & + \mathbf{W}^{[i]}(f) \end{aligned} \quad (9)$$

ここで， $\mathbf{I}$  は単位行列， $\langle \cdot \rangle_t$  は時間平均演算子， $\Phi(\cdot)$  は非線形関数であり，Sawada らによって以下の式によるものが提案されている [18]．

$$\Phi(\mathbf{Y}(f, t)) \equiv \begin{bmatrix} e^{j \arg(Y_1(f, t))} \\ \vdots \\ e^{j \arg(Y_L(f, t))} \end{bmatrix} \quad (10)$$

ここで， $\arg(\cdot)$  は複素数の偏角を求める演算子である．

FDICA では，異なる周波数で独立に学習をすることにより，分離信号の順序の不定性の問題（permutation 問題）と利得の不定性の問題（scaling 問題）が生じる．Permutation 問題に対しては，方位特性に基づく Kurita らの解決法 [16, 22]，また，分離信号の相関と音声の調波構造に基づく Sawada らの解決法 [23] などが提案されている．一方 scaling 問題に対して，Murata らは分離信号のマイクロホン空間への逆射影（Projection Back：PB）による解決法を提案している [12]．

ICA に基づく BSS 手法は，音源の独立性の仮定以外の事前情報を一切必要としないため，音源分離において注目されている技術である．しかしながら，特に残響のような複素畳み込み混合問題を解決する際，非線形最適化の収束問題は遅いという ICA 固有の問題がある [24]．そのうえ，通常 ICA に基づく BSS 手法は，膨大な計算量を必要とする．リアルタイム処理が必要とされる音声を扱うアプリケーションにとって，これは大きな欠点となる．

## 2.4 従来のリアルタイム BSS システム

表 1 にそれぞれの利点・問題点を整理する．BM に基づく BSS は，ごく少量の計算量しか必要としないためリアルタイム実装は比較的容易であるという利点はあるが，スパース性が必要で性能が比較的悪い．一方，ICA は分離性能が高いが，計算量が膨大でリアルタイム実装には不向きである．

表 1 従来の BSS 手法に関する利点・問題点

BSS 手法	利点	問題点
バイナリマスクに基づく BSS	リアルタイム動作に向く	スパース性が必要 分離音が歪む
ICA に基づく BSS	分離性能が高い	計算量が膨大

もちろん，音声を扱うアプリケーションにとって，BSS システムのリアルタイム処理化は必須である．ICA の高い分離性能を生かすために，リアルタイム実装を試みたいいくつかの研究報告例がある [18, 26]．しかし，これら全ての報告は，高速計算機を必要とし，小型 LSI への実装例はない．小型 LSI へのリアルタイム実装，およびそれを用いた小型装置の研究開発は，産業応用へ大きなインパクトがある．

## 2.5 本章のまとめ

本章では，本研究において前提とする混合過程の定式化を行った．また，音源分離の従来法として，バイナリマスクに基づく BSS，ICA に基づく BSS について述べ，さらにこれら手法に関するリアルタイム BSS システムについても述べた．

既に述べたとおり，ICA に基づいた BSS 手法の高性能を維持し，かつ，リアルタイム処理可能な BSS の小型装置の実現は極めて困難である．3 章では，この問題の解決法として，リアルタイム処理向けの新しい BSS 手法を示し，本手法を小型 LSI へリアルタイム実装して開発するリアルタイム音源分離マイクロホンについて詳しく述べる．



## 3. ポケットサイズブラインド音源分離マイクロホン

### 3.1 まえがき

従来の ICA 装置は，2.4 に示す通り，高速計算機を必要とし，小型 LSI へのリアルタイム実装例はない．本章では，まずはじめに，3.2 節で，リアルタイム処理向けの新しい 2 段型 BSS 手法について説明し，さらに，3.3 節で，本手法を小型 LSI へリアルタイム実装して開発するリアルタイム音源分離マイクロホンについて述べる．

さらに，音源分離マイクロホンの基本性能評価として，3.4 節で，雑音抑圧性能，音質性能，それらを統合した主観音質性能の観点での評価実験とその結果について述べる．

加えて，音源分離マイクロホンのリアルタイム処理性能評価として，3.5 節で，移動音源による雑音抑圧性能，ポーラパターンの評価実験とその結果について述べる．

### 3.2 SIMO-ICA と SIMO-BM を統合した 2 段型 BSS

#### 3.2.1 アルゴリズム構成

我々の研究グループでは，ICA の高分離性能を生かしつつリアルタイム処理が可能な 2 段型の BSS 手法を研究してきた [28]．ここでの 2 段型とは，Single-Input Multiple-Output (SIMO) モデルに基づく ICA (SIMO-ICA) と SIMO モデルに基づくバイナリマスク (SIMO-BM) を統合した特徴を示す．

本節では，SIMO-ICA と SIMO モデルに基づくバイナリマスクを結合した提案 2 段型 BSS の構成と特徴について説明する（図 6 参照）．

処理手順は以下の通りである．

前段 観測信号に対し，各音源に関する SIMO モデル信号を出力する SIMO-ICA の一種である，周波数領域 SIMO-ICA (FD-SIMO-ICA) を動作させることにより全音源に関する SIMO モデル信号  $Y_t^{\text{ICA}i}(f, t)$  を得る．ただし，この目的信号には非目的音成分が残留する．

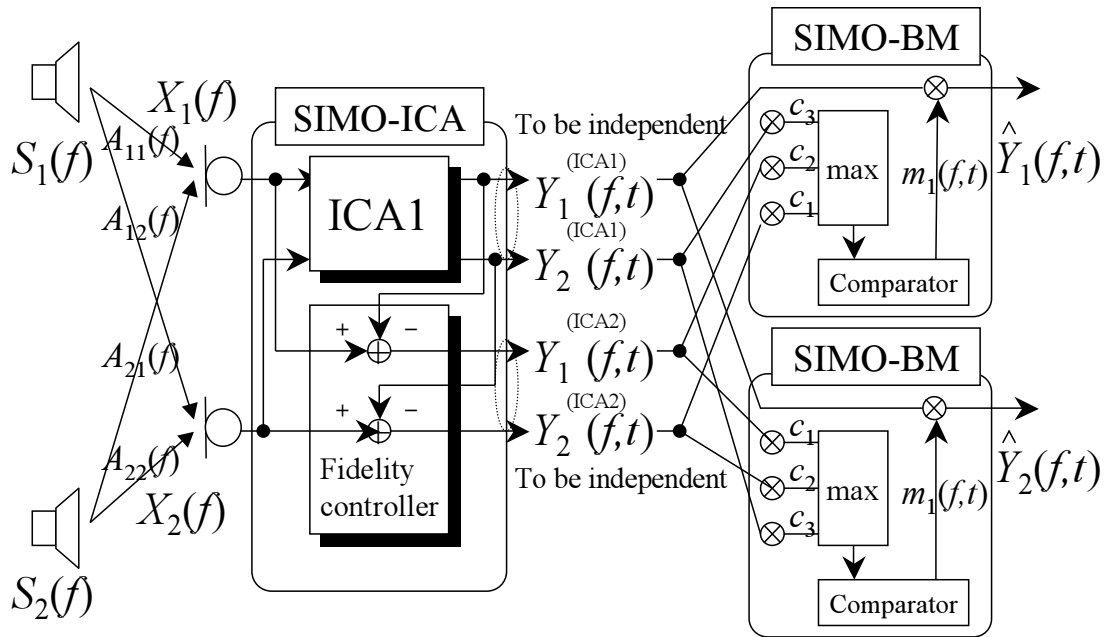


図 6 FD-SIMO-ICA と SIMO-BM を接続した提案 2 段 BSS :  $K = L = 2$  の場合  
 後段 得られた SIMO モデル信号に対し、後処理として SIMO モデルに基づくバイナリマスクを用いることにより FD-SIMO-ICA の出力信号に含まれる残留成分の除去を行う。

SIMO-ICA とは、混合信号を、モノラル信号ではなく SIMO モデル信号と呼ばれるマイクロホン観測時点でのアレー信号に分離することができる特徴を持つ ICA 手法である [13, 20, 21]。SIMO-ICA は、従来の ICA に基づく BSS 手法と比較して、各音源の空間的特性を保持したままの分離信号を得ることができる。SIMO-ICA の出力は、推定 (分離) された SIMO モデルに基づく信号であり、一組のアレー信号となっている。マイク素子が指向特性を持つ場合、 $l$  番目のマイクロホン中の最も支配的な信号は  $l$  番目の音源信号成分であるという関係が保たれている。

SIMO-BM とは、図 6 に示すように、前述の SIMO モデルに基づくアレー信号を入力として、2.3 章で説明したバイナリマスク手法を拡張して構成している。

ICA とバイナリマスクの接続は，SIMO-ICA に拡張しないモノラル出力 ICA とバイナリマスクの単純接続 [30] が容易に考えられるが，SIMO モデルに基づくアレー信号を入力を考慮することで，分離対象とする音源信号間の音量差に影響を受けず，スパース性の仮定が全く成立しない場合（分離対象とする音源信号が存在する周波数帯域が重なる等）においても，ICA の残留誤差を効率的に削減することが可能である．これは，聴覚的にも妥当な処理と考えられる．

次節では，提案法の前段において SIMO モデル信号を出力する，FD-SIMO-ICA について詳しく説明を行う．

### 3.2.2 周波数領域 SIMO-ICA (FD-SIMO-ICA)

本節では，周波数領域 SIMO-ICA (FD-SIMO-ICA) の説明を行う．FD-SIMO-ICA は各音源信号に対応する SIMO モデルに基づく信号を求める手法である．FD-SIMO-ICA は  $(L - 1)$  個の FDICA と単一の *fidelity controller* (FC) から成り，各 ICA は分離システム全体の再現精度を保ちながら並列に動作する（図 6 参照）．FD-SIMO-ICA における  $l$  番目の ICA ( $l = 1, \dots, L - 1$ ) は以下で定義される．

$$\mathbf{Y}_{(\text{ICAL})}(f, t) = \mathbf{W}_{(\text{ICAL})}(f) \mathbf{X}(f, t) \quad (11)$$

ここで  $\mathbf{W}_{(\text{ICAL})}(f) = [W_{ij}^{(\text{ICAL})}(f)]_{ij}$  は， $l$  番目の FDICA における分離フィルタ行列である．

また，FC は以下の式で表される信号を出力する．

$$\mathbf{Y}_{(\text{ICAL})}(f, t) = \mathbf{X}(f, t) - \sum_{l=1}^{L-1} \mathbf{Y}_{(\text{ICAL})}(f, t) \quad (12)$$

以後， $\mathbf{Y}_{(\text{ICAL})}(f, t)$  を，仮想的な  $L$  番目の ICA の出力とみなす．ここで「仮想的な」という単語を使う訳は， $L$  番目の ICA はその他の ICA と違い独自の分離フィルタを持たず， $\mathbf{Y}_{(\text{ICAL})}(f, t)$  は  $\mathbf{W}_{(\text{ICAL})}(f)$  ( $l = 1, \dots, L - 1$ ) に従属しているためである．右辺の第二項 ( $-\sum_{l=1}^{L-1} \mathbf{Y}_{(\text{ICAL})}(f, t)$ ) を左辺に移項すると，式 (12) は，全 ICA の出力ベクトルの和  $\sum_{l=1}^L \mathbf{Y}_{(\text{ICAL})}(f, t)$  が，全 SIMO 成分の

和  $[\sum_{l=1}^L A_{kl}(f)S_l(f,t)]_{k1}(= \mathbf{X}(f,t))$  になるための拘束条件になっていることが分かる .

式 (11) により独立な音源が分離され , 同時に 式 (12) により得られた信号が互いに独立であれば , 出力信号は次のような一意な SIMO モデル信号に収束する .

$$\mathbf{Y}_{(\text{ICA}l)}(f,t) = \text{diag}[\mathbf{A}(f)\mathbf{P}_l^T] \mathbf{P}_l \mathbf{S}(f,t) + \mathbf{E}_l(f,t) \quad (13)$$

ここで  $\text{diag}[\mathbf{X}]$  は行列  $\mathbf{X}$  の対角要素以外を全て 0 で置き換える操作であり ,  $\mathbf{P}_l$  ( $l = 1, \dots, L$ ) は  $\sum_{l=1}^L \mathbf{P}_l = [\mathbf{1}]_{ij}$  となるような排他的置換行列群 [25] である . この証明は [13] を周波数領域に拡張することにより得られる . 式 (13) により与えられる解は , 各  $l$  番目の音源に関する必要十分な SIMO 成分 ,  $A_{kl}(f)S_l(f,t)$  , を与えることは明白である . そのため , SIMO-ICA の分離信号は , 各音源信号の空間特性を保つことが可能である .  $L = K = 2$  の場合 , ICA の出力は次の式で与えられる .

$$\begin{aligned} & \left[ Y_1^{(\text{ICA}1)}(f,t), Y_2^{(\text{ICA}1)}(f,t) \right]^T \\ & = [A_{11}(f)S_1(f,t) + E_{11}(f,t), A_{22}(f)S_2(f,t) + E_{22}(f,t)]^T, \quad (14) \end{aligned}$$

$$\begin{aligned} & \left[ Y_1^{(\text{ICA}2)}(f,t), Y_2^{(\text{ICA}2)}(f,t) \right]^T \\ & = [A_{12}(f)S_2(f,t) + E_{12}(f,t), A_{21}(f)S_1(f,t) + E_{21}(f,t)]^T, \quad (15) \end{aligned}$$

ここで ,  $\mathbf{P}_1 = \mathbf{I}$  ,  $\mathbf{P}_2 = [\mathbf{1}]_{ij} - \mathbf{I}$  としている .

式 (13) を得るため , 式 (12) の Kullback-Leibler Divergence の  $\mathbf{W}_{(\text{ICA}l)}(f)$  に関する Natural Gradient を ,  $l$  番目 ( $l = 1, \dots, L - 1$ ) の ICA における分離フィルタの Non-Holonomic 反復学習式 [12] に追加する必要がある . したがって , FD-SIMO-ICA における  $l$  番目 ( $l = 1, \dots, L - 1$ ) の ICA 部の新しい反復学習

アルゴリズムは以下で与えられる .

$$\begin{aligned}
& \mathbf{W}_{(\text{ICA}l)}^{[j+1]}(f) \\
&= \mathbf{W}_{(\text{ICA}l)}^{[j]}(f) - \alpha \text{ off-diag} \left\langle \Phi \left( \mathbf{Y}_{(\text{ICA}l)}^{[j]}(f, t) \right) \mathbf{Y}_{(\text{ICA}l)}^{[j]}(f, t)^{\text{H}} \right\rangle_t \mathbf{W}_{(\text{ICA}l)}^{[j]}(f) \\
&+ \alpha \text{ off-diag} \left\langle \Phi \left( \mathbf{X}(f, t) - \sum_{\nu=1}^{L-1} \mathbf{Y}_{(\text{ICA}\nu)}^{[j]}(f, t) \right) \left( \mathbf{X}(f, t) - \sum_{\nu=1}^{L-1} \mathbf{Y}_{(\text{ICA}\nu)}^{[j]}(f, t) \right)^{\text{H}} \right\rangle_t \\
&\quad \left( \mathbf{I} - \sum_{\nu=1}^{L-1} \mathbf{W}_{(\text{ICA}\nu)}^{[j]}(f) \right) \tag{16}
\end{aligned}$$

ここで ,  $\alpha$  は更新係数で , 非線形ベクトル関数  $\Phi(\cdot)$  は次のように定義する [18] .

$$\begin{aligned}
\Phi(\mathbf{Y}(f, t)) \equiv & \left[ \tanh(|Y_1(f, t)|) e^{j \arg(Y_1(f, t))}, \dots, \right. \\
& \left. \tanh(|Y_L(f, t)|) e^{j \arg(Y_L(f, t))} \right]^{\text{T}}. \tag{17}
\end{aligned}$$

また ,  $\mathbf{W}_{(\text{ICA}l)}(f)$  の初期値は全て異なっている必要がある .

### 3.2.3 SIMO モデルに基づくバイナリマスク

提案法の後段では , SIMO モデルに基づくバイナリマスク (SIMO-BM) を用いる (図 6 参照) . 今回は , FD-SIMO-ICA において式 (14) , (15) の場合を考える . 提案 SIMO-BM における音源 1 に対する出力信号は以下の式で表すことができる .

$$\hat{Y}_1(f, t) = m_1(f, t) Y_1^{(\text{ICA}1)}(f, t) \tag{18}$$

ここで ,  $m_1(f, t)$  は SIMO-BM 演算子で , 次の条件のときに  $m_1(f, t) = 1$  となり , それ以外の場合は 0 となる .

$$Y_1^{(\text{ICA}1)}(f, t) > \max \left[ \left| c_1 Y_2^{(\text{ICA}2)}(f, t) \right|, \left| c_2 Y_1^{(\text{ICA}2)}(f, t) \right|, \left| c_3 Y_2^{(\text{ICA}1)}(f, t) \right| \right] \tag{19}$$

ここで  $\max[\cdot]$  は引数の中での最大値をとる関数を表し ,  $c_1, \dots, c_3$  はマスクを生成する際の各 SIMO 成分の重みである . 例えば  $[c_1, c_2, c_3] = [0, 0, 1]$  は , 従来の ICA と従来のバイナリマスク処理との単純接合 [30] と等価である . また ,  $[c_1, c_2, c_3] =$

[1, 0, 0] と設定した場合，3.2.1 節にて述べた，聴覚的に妥当な，各音源に関する SIMO 情報を活用することができる．このように  $c_i$  の値のパターンを変更することで様々な SIMO-BM を表現することができ，分離性能や歪の度合いを調整可能である．

音源 2 に対応する出力は以下の式で与えられる．

$$\hat{Y}_2(f, t) = m_2(f, t) Y_2^{(\text{ICA1})}(f, t) \quad (20)$$

ここで， $m_2(f, t)$  は SIMO-BM 演算子で，次の条件のときに  $m_2(f, t) = 1$  となり，それ以外の場合は 0 となる．

$$Y_2^{(\text{ICA1})}(f, t) > \max \left[ \left| c_1 Y_1^{(\text{ICA2})}(f, t) \right|, \left| c_2 Y_2^{(\text{ICA2})}(f, t) \right|, \left| c_3 Y_1^{(\text{ICA1})}(f, t) \right| \right] \quad (21)$$

一般系である  $L = K > 2$  に拡張した場合も，簡単に式を与えることができる．以降では，排他的置換行列が以下の場合を一例として考える．

$$\mathbf{P}_l = [\delta_{in(k,l)}]_{ki} \quad (22)$$

ここで， $\delta_{ij}$  はクロネッカーのデルタ関数を表し，

$$n(k, l) = \begin{cases} k + l - 1 & (k + l - 1 \leq L) \\ k + l - 1 - L & (k + l - 1 > L) \end{cases} \quad (23)$$

である．この場合，式 (13) は以下のように評価される．

$$\mathbf{Y}_{(\text{ICAL})}(f, t) = [A_{kn(k,l)}(f) S_{n(k,l)}(f, t) + E_{kn(k,l)}(f, t)]_{k1} \quad (24)$$

このように，SIMO-BM における音源 1 に対する出力は以下の式で与えられる．

$$\hat{Y}_1(f, t) = m_1(f, t) Y_1^{(\text{ICA1})}(f, t) \quad (25)$$

ここで， $m_1(f, t)$  は以下の条件のとき 1 と定義され，それ以外の場合は 0 となる．

$$Y_1^{(\text{ICA1})}(f, t) > \max \left[ \left| c_1 Y_2^{(\text{ICAL})}(f, t) \right|, \left| c_2 Y_3^{(\text{ICAL-1})}(f, t) \right|, \left| c_3 Y_4^{(\text{ICAL-2})}(f, t) \right|, \right. \\ \left. \dots, \left| c_{L-1} Y_L^{(\text{ICA2})}(f, t) \right|, \dots, \left| c_{LL-1} Y_L^{(\text{ICA1})}(f, t) \right| \right] \quad (26)$$

これ以外の音源についても，同様の操作で出力信号を得ることができる．

### 3.2.4 本構成の特徴

本節で本構成の特徴をまとめる．この 2 段型 BSS の優位性は，SIMO-ICA と SIMO-BM の接続方法にある．前段の SIMO モデルに基づく分離信号は一組のアレー信号となっていることから，後段に SIMO モデルに基づく複数チャンネル信号処理を結合することで，新たな効果を生み出す．まず，マイクロホンを指向性とすることにより，SIMO-ICA の出力は，例えば  $l$  番目のマイクロホンは  $l$  番目の音源信号をより強く観測する，といった特性を有する．SIMO-ICA の後段に接続されている SIMO-BM は，一つの音源のアレー信号同士でその信号強度比較をするので，分離対象の複数の音源同士にスパース性の仮定が全く成立しない場合においても，ICA の残留誤差を効率的に削減することが可能である．よって，ICA 単独の場合よりも分離性能が高い．

さらに，SIMO-BM は少量の計算量でリアルタイム動作可能なので，SIMO-ICA が分離フィルタ更新に時間を要し，リアルタイム処理できない場合でも SIMO-BM がそれをカバーして，全体としてリアルタイム動作可能となる．この組み合わせの考え方は，リアルタイム実装にも向いている．

このように，本構成は，分離性能が高く，かつリアルタイム動作可能な構成といえる．

## 3.3 リアルタイム BSS マイクロホン

### 3.3.1 概要

我々は，雑音環境下でリアルタイムに音源分離が可能な，SSM-001 と名づけた新しい音源分離マイクロホンを開発した [39]．図 7 は初期のプロトタイプ [38]，図 8 に音源分離マイクロホン SSM-001 示す．主な仕様は，表 2 に示す．ハードウェアブロック図は，図 9，内部基板は図 10 に示す．なお，本マイクロホンはハードウェアとしては 3 つのマイク素子まで取り扱うことが可能だが，本章でのリアルタイム実装は左右 2 つのマイク素子を用いて 2 入力システムとする．また，それぞれのマイク素子は単一指向性を持っている．

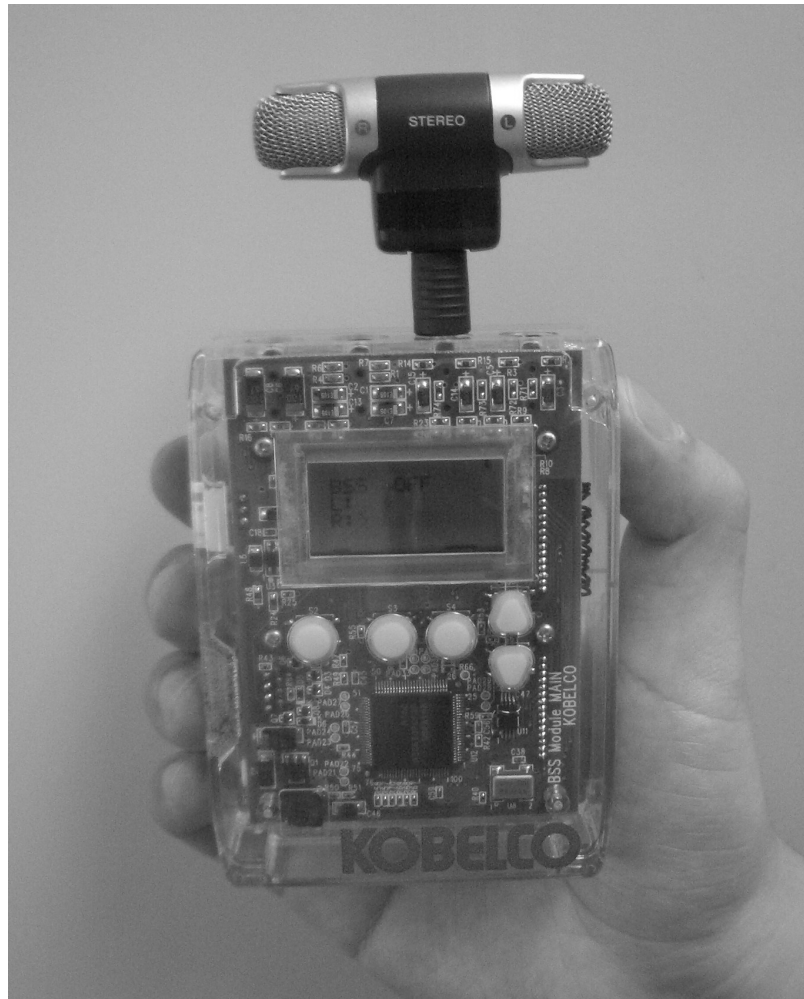


図 7 BSS モジュール(1号試作機)





図 8 BSS マイクロホン (SSM-001)

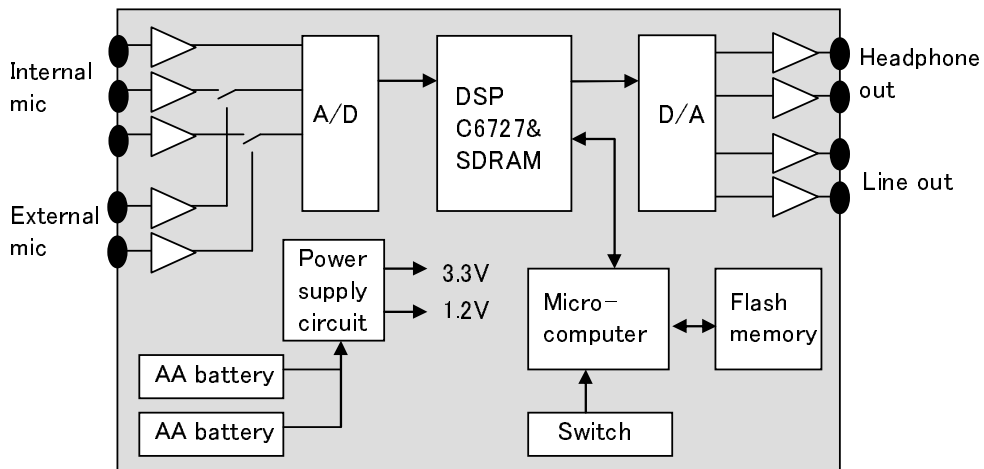


図 9 BSS マイクロホンのハードウェア構成

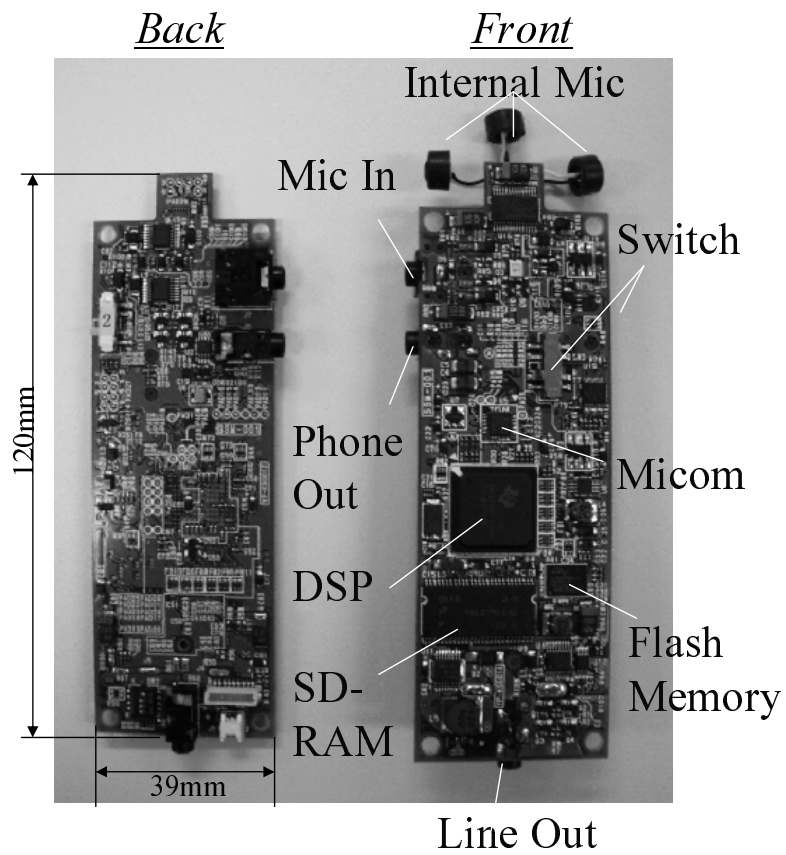


図 10 BSS マイクロホンの内部基板

表 2 BSS マイクロホン諸元概要

processor	TI-DSP TMS320VC6727 (Clock 300 MHz)
input	internal stereo mic
	external stereo mic (plug-in power)
output	line out
	headphone out
terminal	external mic in
	line out
	headphone out
sampling frequency	16 kHz or 8 kHz
battery	AA battery × 2
memory	Flash ROM: 8 MB (used about 330 KB)
	SDRAM: 128 MB (used about 1.9 MB)
size	136 mm (H) × 45 mm (W) × 27 mm (D)
weight	125 g (including battery)

見てのとおり，SSM-001 はポケットサイズにまで小さくした今までに前例のない世界最小のリアルタイム BSS 装置である。

### 3.3.2 リアルタイム実装

図 11 に 2 段型 BSS アルゴリズムのリアルタイム実装に関連して処理フローの説明図を示す。また，図 12 に ソフトウェア処理の観点から，DSP および SDRAM の内部ブロック図を示す。リアルタイム実装の信号処理フローは次の処理の流れ通りである。

[Step 1] 入力信号は高速フーリエ変換 (FFT) により，時間周波数信号列  $X(f, t)$  にフレーム単位に変換させる。この時の DSP 内部での処理は，次の通りである。DSP が持つマルチチャンネルオーディオシリアルポート

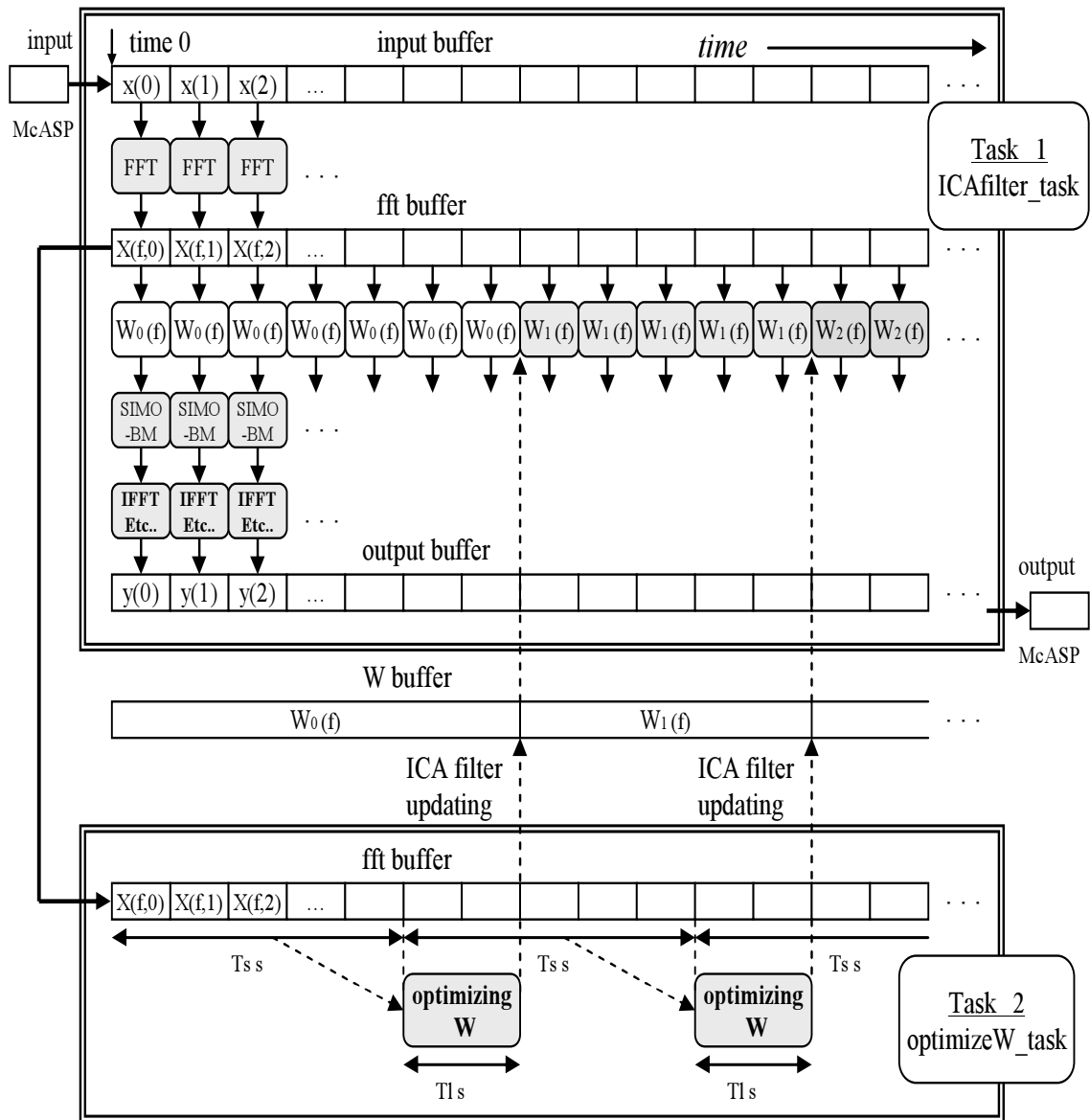


図 11 リアルタイム処理の流れ

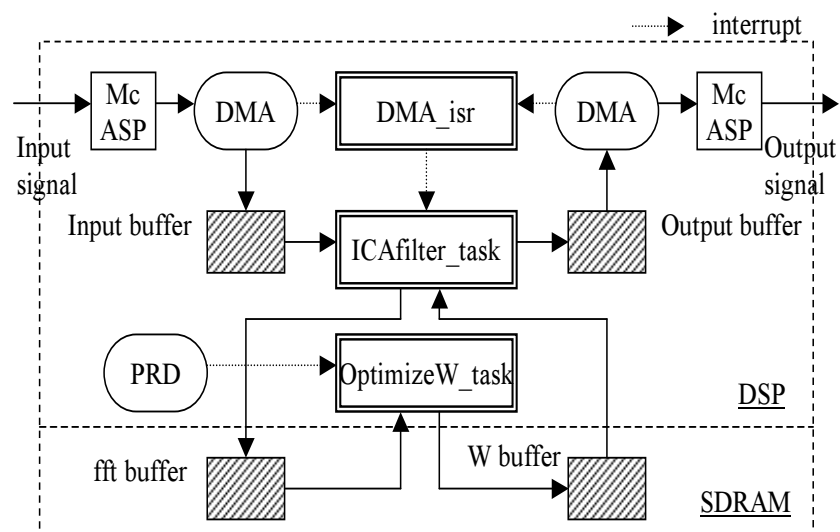


図 12 DSP と SDRAM の内部ブロック構成

(*multichannel audio serial port: McASP*) を通じて、左右それぞれの各マイクロホンからの入力をサンプリングレートで指定された時間単位毎に取り込み、(*direct memory access: DMA*) 機能により入力バッファに送られる。入力バッファが一杯になると、DMA は割り込みを発生し、DMA 割り込みルーチン (*DMA\_isr*) を起動する。DMA 割り込みルーチンでは、*ICA\_filter\_task* が呼び出され、*ICA\_filter\_task* 内で、入力バッファに貯められた入力信号は、時間周波数信号列に変換される。

[Step 2] SIMO-ICA は (図 11 の  $T_s$ ) 秒毎の入力信号を用いて分離行列の推定を行う。その分離行列は最適化が終了次第すぐに更新され、更新後からの入力信号に適用される。SIMO-ICA 中の ICA 更新式は、次の式で与えられる。

$$\begin{aligned}
& \mathbf{W}_{(\text{ICA1})}^{[j+1]}(f) \\
&= \mathbf{W}_{(\text{ICA1})}^{[j]}(f) - \alpha \left[ \left\{ \text{off-diag} \left\langle \Phi(\mathbf{Y}_{(\text{ICA1})}^{[j]}(f, t)) \right. \right. \right. \\
&\quad \left. \left. \left. \mathbf{Y}_{(\text{ICA1})}^{[j]}(f, t)^{\text{H}} \right\rangle_t \right\} \cdot \mathbf{W}_{(\text{ICA1})}^{[j]}(f) \right. \\
&\quad \left. - \left\{ \text{off-diag} \left\langle \Phi(\mathbf{X}(f, t) - \mathbf{Y}_{(\text{ICA1})}^{[j]}(f, t)) \right. \right. \right. \\
&\quad \left. \left. \left. \cdot (\mathbf{X}(f, t) - \mathbf{Y}_{(\text{ICA1})}^{[j]}(f, t))^{\text{H}} \right\rangle_t \right\} \right. \\
&\quad \left. \cdot (\mathbf{I} - \mathbf{W}_{(\text{ICA1})}^{[j]}(f)) \right], \tag{27}
\end{aligned}$$

ここで,  $\alpha$  はステップサイズパラメータ,  $\Phi(\cdot)$  は適当な非線形関数である .  
ここで, パーミュテーションの影響を考慮することなく ([13] の証明を参照),  
SIMO-ICA を通じて  $\mathbf{Y}_{(\text{ICA}l)}(f, t) = [Y_1^{(\text{ICA}l)}(f, t), Y_2^{(\text{ICA}l)}(f, t)]^{\text{T}}$  ( $l = 1, 2$ )  
の解を得ることができる .

$$\begin{aligned}
& \mathbf{Y}_{(\text{ICA1})}(f, t) \\
&= \mathbf{W}_{(\text{ICA1})}(f) \mathbf{X}(f, t) \\
&= [A_{11}(f)S_1(f, t), A_{22}(f)S_2(f, t)]^{\text{T}} + \mathbf{E}_1, \tag{28}
\end{aligned}$$

$$\begin{aligned}
& \mathbf{Y}_{(\text{ICA2})}(f, t) \\
&= (\mathbf{I} - \mathbf{W}_{(\text{ICA1})}(f)) \mathbf{X}(f, t) \\
&= [A_{12}(f)S_2(f, t), A_{21}(f)S_1(f, t)]^{\text{T}} + \mathbf{E}_2, \tag{29}
\end{aligned}$$

ここで,  $\mathbf{E}_i$  は消し残り誤差である . SIMO-ICA の各出力は, 再び各マイク  
位置でのアレー信号になっているため, SIMO-ICA 処理後段にバイナリマス  
ク処理を繋いでこの消し残り誤差  $\mathbf{E}_i$  を消す . この時, DSP (図 12) 内部処  
理は次の通りである . SIMO-ICA のフィルタ更新は, 独立したタスク *Opti-*  
*mize W\_task*内で行われる . この *Optimize W\_task*は, (*PRD:periodic function*

*manager*) によって定期的に呼び出され, 時間周波数信号列が貯められる *fft buffer* をチェックする. *fft buffer* が一杯になっていれば, *Optimize W\_task* は分離行列  $W(f)$  の最適化を行い, *W buffer* に格納されている分離行列  $W(f)$  の更新を行う. SIMO-ICA のフィルタ更新は計算が非常に複雑であるが, 3 秒のデータを用いて 100 回の繰り返し最適化計算を約 0.5 秒で実施するレベルまで実現できている (図 11 中の  $T_l$  に相当). そのため, 3 秒データを用いた場合のトータルのフィルタの更新遅れは, 3 秒 + 0.5 秒の 3.5 秒である.

[Step 3] 前段の SIMO-ICA 処理で得られた分離信号を用いて SIMO-BM 処理が行われる. SIMO-ICA 処理とは異なり, バイナリマスク処理は現在の時間の信号に対して処理可能である. 音源 1 に相当する出力信号は, SIMO-BM 処理内で次のように計算される.

$$\hat{Y}_1(f, t) = m_1(f, t)Y_1^{(\text{ICA1})}(f, t), \quad (30)$$

ここで,  $m_1(f, t)$  は, SIMO モデルに基づく バイナリマスクパラメータで, 次式で定義される.

$m_1(f, t) = 1$  if

$$Y_1^{(\text{ICA1})}(f, t) > \max [|c_1 Y_2^{(\text{ICA2})}(f, t)|, |c_2 Y_1^{(\text{ICA2})}(f, t)|, |c_3 Y_2^{(\text{ICA1})}(f, t)|]; \quad (31)$$

otherwise,  $m_1(f, t) = 0$ .

ここで,  $c_i$  は分離性能と音質のバランスを調整するパラメータで, 通常は,  $c_1 = 1$ ,  $c_2 = 0$ , and  $c_3 = 0 \sim 1$  を用いる. 音源 2 についても同様に求められる.

[Step 4] SIMO-BM の出力信号は, 逆高速フーリエ変換 (IFFT) により, 時間領域の波形信号に変換される. この時, SIMO-BM と IFFT 処理は図 12 中の *ICA\_filter\_task* にて行われる.

SIMO-ICA のフィルタ更新はリアルタイム処理ではないため前述の通り数秒の更新遅れが存在するが，SIMO-BM が現在の時間信号に対して瞬時処理可能なため，全体の 2 段型処理はリアルタイムで動作する．一般に，ICA はフィルタの更新遅れがあるために問題となり，リアルタイム処理を必要とするアプリケーションへの応用が難しかった．しかし，本手法では，SIMO-ICA の遅延による分離性能劣化はバイナリマスクの効果により大きく軽減できている．

### 3.3.3 遅延時間

入力信号に対応する出力信号を得るまでの時間を「遅延時間」と呼ぶ．以下に，音源分離マイクロホンでの遅延時間について説明するために，DSP 内部でのデータの流りに焦点を置いて説明する．

まず，全体の遅延時間のイメージを図 13 を用いて説明する．左右それぞれのマイクロホン毎にバッファサイズ  $N$  を単位として音源分離処理（図 13 の Proposed BSS）に渡される．①は初期化状態である．図中の Input Buffer の Buffer1 の先頭位置を受信開始ポイントとしてデータを受信する．また，Output Buffer の Buffer1 の先頭位置を送信開始ポイントとしてデータを送信する．これら 2 つのポイントはサンプリング周期に同期しながら進む．②は Input Buffer の Buffer1 が入力データで一杯になった直後である．受信したデータを音源分離処理に渡す．③は Buffer2 に受信を開始してしばらく経過したところである．音源分離処理が終了して Output Buffer の Buffer1 にデータが出力がされている状態を示す．④は Input Buffer の Buffer2 が一杯になり，音源分離処理にデータを渡したところである．⑤はバッファを一巡して再度入力バッファの第一バッファに受信しているところである．この時，音源分離処理からの最初の出力も同時に出力されている．この入出力バッファの管理により， $2N$  サンプルの遅延となる．

しかし，詳細は後に述べるが，音源分離処理においても，オーバーラップ処理で現時刻より前の信号との重ねあわせによる平滑化処理を行っているため， $N$  サンプルの遅延がある．よって，入出力のバッファ管理による  $2N$  サンプルの遅延と音源分離処理での  $N$  サンプルの遅延をあわせた  $3N$  サンプルの遅延が発生することになる．なお，入出力のバッファ管理の遅延は音源分離処理での処理時間を



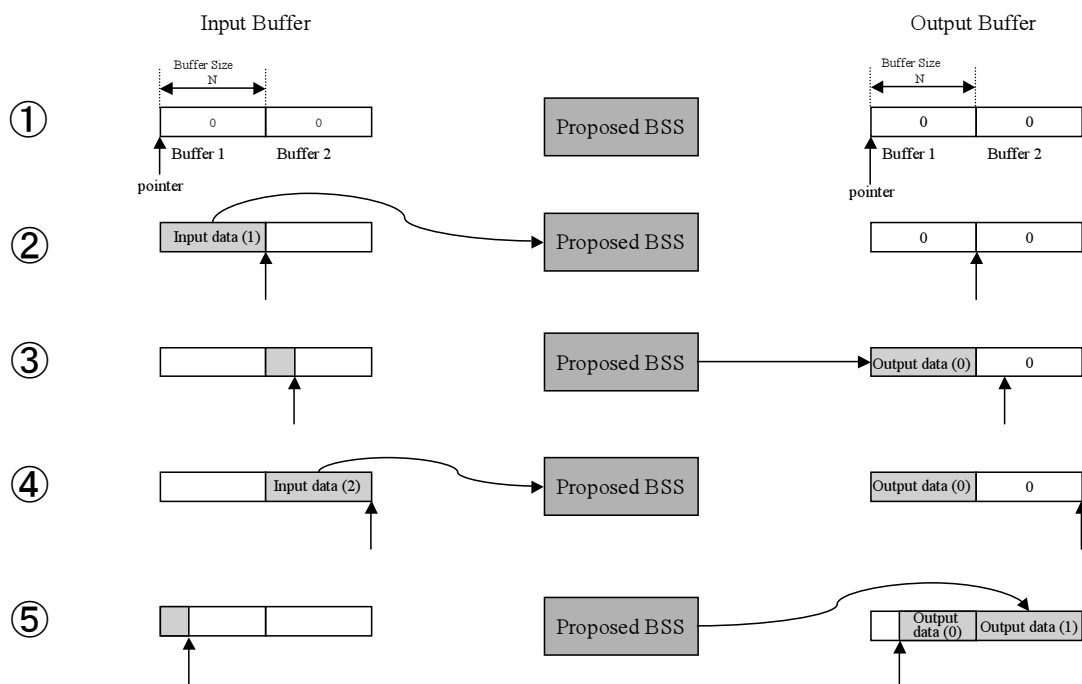


図 13 入出力バッファの管理

考慮して、処理完了次第出力を開始するなど送信開始ポイントを予め前に設定することで短縮も可能であるため、処理遅延時間は  $3N$  サンプルより短くすることも可能である。

次に、音源分離処理の内部についても、遅延時間の観点からもう少し詳しく説明する。3.3.2 節の Step1 において、入力信号は高速フーリエ変換 (FFT) により、時間周波数信号列にフレーム単位に変換させるが、この時、音源分離性能の観点から、8 kHz サンプルングにおいては、1024 サンプルのフレーム長が最もよいとされている [40]。16 kHz の場合には、周波数分解能を維持するとすれば、2048 サンプルのフレーム長となる。このフレーム長は時間に直せば 128 ms となり、上述の処理遅延時間は  $2048 \times 3$  サンプル分、すなわち  $128 \text{ ms} \times 3 = 384 \text{ ms}$  となる。しかし、デジタル携帯電話での遅延時間は一般に 50 msec 以下が普通であり、音源分離処理を携帯電話へ応用する場合には、さらに大幅な遅延時間増となり利便性に支障をきたす。さらに補聴器への応用では対話相手などの目で見える映像と

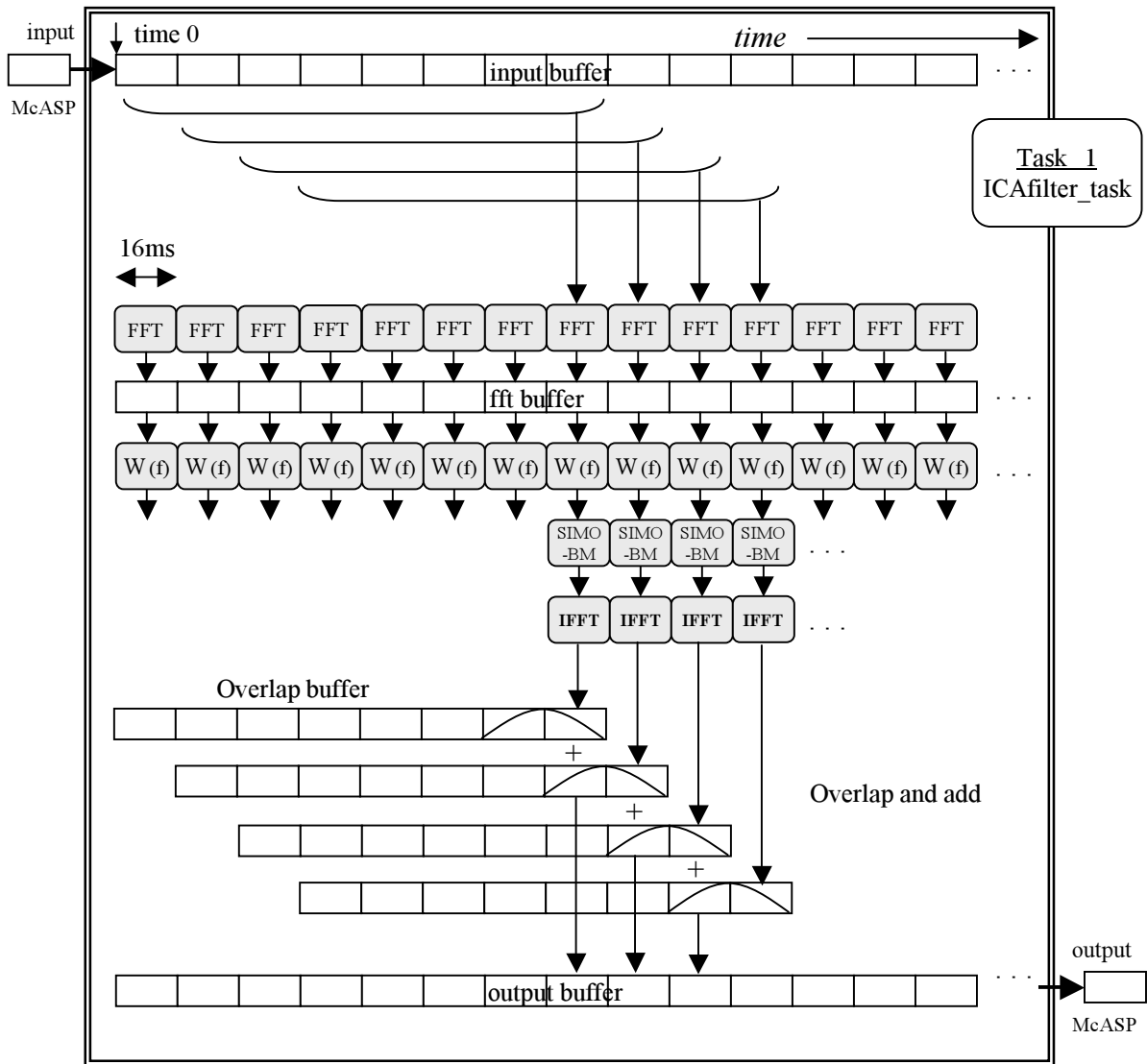


図 14 遅延時間短縮のためのオーバーラップ加算 (8 倍の場合)

表 3 提案法リアルタイム実装時のリソース

ROM 量	330 kByte
RAM 量	1.86 Mbyte
処理量 (CPU 占有率)	46 % (138/300 MHz) @ 16 kHz フィルタ更新処理無の場合 (有の場合は 100 %)

実際に聞こえる音の感覚的なズレが無視できなくなり、実用的ではない。よって、本研究では、多くのアプリケーションへの応用に少しでも近づけるために、DSP リソースの許す範囲で図 13 のバッファサイズ  $N$  を極力小さくし、 $N$  サンプルと入力済の過去のサンプルによって 2048 サンプルフレーム生成して、FFT 処理による周波数分解能を確保する。現状の DSP では、図 14 に示すように、 $N = 256$  サンプルまで小さくすることができ、48 ms の遅延時間を実現している。DSP の能力次第で遅延時間はさらに小さくできる。

### 3.3.4 リソース

前節で説明したリアルタイム実装において、各リソースを表 3 にまとめる。C 言語によるソースプログラムを開発し、DSP 開発の統合環境である CCS ver.3 で標準モードでのコンパイルを行う。処理中最も負荷が高くなるケースは、バックグラウンドで動作するフィルタ更新処理が実施される場合であり、この時 MPU の占有率はほぼ 100%に達する。

## 3.4 基本性能評価

### 3.4.1 評価値

BSS マイクロホンの基本性能を把握するために、次の 3 つの評価値で評価する。 $NRR$  (noise reduction rate),  $CD$  (cepstral distortion),  $PESQ$  MOS-LQO の 3 つの評価値である。

### 3.4.1.1 Noise Reduction Rate (NRR)

Noise reduction rate (NRR) [16] は出力信号の signal-to-noise ratio (SNR) – 入力信号 SNR で定義され、出力信号の分離性能を評価する。SNR は非目的話者の音声を noise として算出する。NRR は以下の式で与えられる。

$$\text{NRR [dB]} \equiv \frac{1}{2} \sum_{l=1}^L (\text{OSNR}_l - \text{ISNR}_l) \quad (32)$$

ここで、 $\text{OSNR}_l$  と  $\text{ISNR}_l$  はそれぞれ、音源  $l$  に関する出力と入力の SNR を表す。分離が完全に達成されたとき、この評価値は無限大を示す。

NRR は分離信号の歪みについて考慮を行っていないため、3.4.1.2 目に述べる cepstral distortion と併用して評価を行う必要がある。

### 3.4.1.2 Cepstral Distortion (CD)

Cepstral distortion (CD) [35] は出力信号中の音源信号成分と音源信号間のケプストラム上での距離で定義される評価値で、出力信号の歪みを評価する。CD は以下の式で与えられる。

$$\text{CD [dB]} \equiv \frac{1}{F} \sum_{f=1}^F D_b \sqrt{\sum_{i=1}^p 2 (C_{\text{out}}(i; f) - C_{\text{ref}}(i; f))^2} \quad (33)$$

ここで、 $F$  は信号のフレーム長、 $C_{\text{out}}(i; f)$  は出力信号中の音源信号成分の  $f$  番目のフレームにおける FFT を用いた  $i$  次のケプストラム係数、 $C_{\text{ref}}(i; f)$  は残響成分が畳み込まれた音声信号（以降 REF 音と称する）のケプストラム係数を表す。また、 $D_b$  は距離尺度をデシベルに変換するための定数で  $D_b = 20/\log 10$ 、 $p$  は評価に用いるケプストラム次数を決定するための値で、本研究では 8 を用いる。出力信号の歪みが少ないほど、この評価値は小さな値を示す。

CD は分離性能について考慮を行っていないため、前述した Noise Reduction Rate (NRR) とは相補的な評価値である。

### 3.4.1.3 PESQ MOS-LQO

PESQ MOS-LQO は、ITU-T Recommendation P.862 で勧告されている音声品質の評価指標 Perceptual Evaluation of Speech Quality (PESQ) の一種 [36] で、最新の客観的手法の一つである。MOS-LQO は Mean Opinion Score - Listening-only Quality for Objective の略である。PESQ MOS-LQO の評価段階は 0.5 ~ 4.5 で 0.5 が最も良い。ITU-T Recommendation P.862.1 で勧告されている。MOS は最も人間の感覚に近く評価できる方法として知られているが、評価者を集める等の手間の問題から、PESQ MOS-LQO はこれと直接比較可能な数式での評価手法として考案された評価値である。これら性質から、NRR と CD 両方に関係し、本研究では、これらの総合評価指標として評価に採用する。

### 3.4.2 実験条件

実験用音声として、2人の男性と2人の女性の合計4人の話者により発声された、研究用 ASJ 音声コーパス [37] 2種類の文章を用いる。これらを用いることにより、 $(\theta_1, \theta_2)$  それぞれにおいて、12種類の実験ペアを得る。サンプリングレートは 8 kHz とする。2つの音声信号は、異なる方向から来ると仮定し、 $(\theta_1, \theta_2) = (0^\circ, \dots, 90^\circ, -90^\circ, \dots, 0^\circ)$  および  $(\theta_1, \theta_2) = (-90^\circ, \dots, 0^\circ, 0^\circ, \dots, 90^\circ)$ 、として、それぞれ  $10^\circ$  刻みで評価する。また、残響時間 200 ms の部屋で集音した多数のインパルス応答を用いて、ドライ音声信号に畳み込んで使用し、混合音をシミュレートする。また、この実験はノイズフリーである。すなわち、加法性の背景雑音は考慮しない。さらに ICA 部における初期行列は、 $(-60^\circ, 60^\circ)$  方位への死角ビームフォーマ [16] を用いる。

### 3.4.3 実験結果

図 15 は、(a)NRR, (b)CD, (c)PESQ MOS-LQO のそれぞれ計測結果である。これら結果は、 $\theta_1$  と  $\theta_2$  の全ての組み合わせの平均値である。比較した手法は (A) 従来のバイナリマスク, (B) 従来の SIMO-ICA, (C) 実装した 2 段型 BSS 法, の 3 つである。(B) および (C) の ICA 部では、分離行列の推定に 3 秒毎のデータを

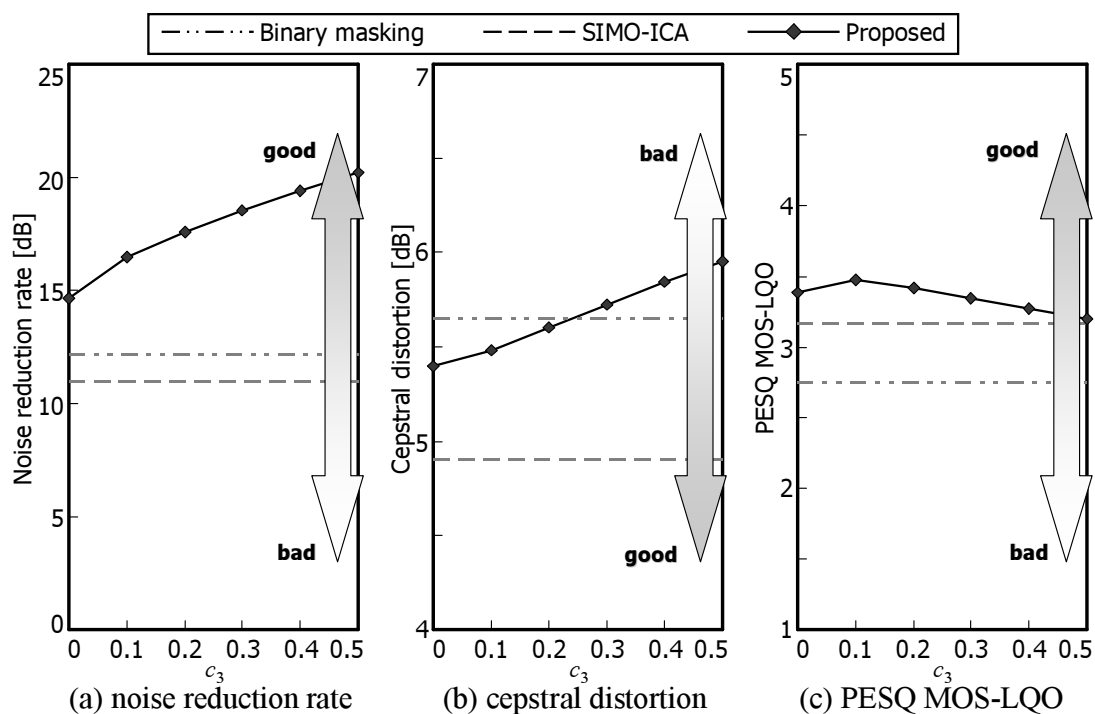


図 15 実験結果:(a) 雑音抑圧量, (b) ケプストラム歪, (c) PESQ MOS-LQO

用いた．この結果より， $c_3$  の増加により，NRR は改善し，CD は大きくなる（音質は劣化する）ことが確認できる．前述したようなトレードオフは存在するが，PESQ MOS-LQO は  $c_3 = 0.1$  が最も良いパラメータである．これは， $c_3$  のゲインを少し与えることで，人間の聴覚的には最も良い性能を与えることができることを意味する．また，より経験的な評価に基づいて言えば，背景雑音が多い環境では， $c_3$  は 0.1 より大きい値に設定する方が良い性能を得ることができる．

以上は客観的評価指標を用いて，提案法ならびに音源分離マイクロホン SSM-001 を評価したが，筆者の主観的評価においても，ほぼ PESQ-MOS LQO 値に沿う傾向が確認されている．

表 4 音源 2( $\theta_2$ ) の移動条件

$\theta_2$	0 → 10 s	10 → 11 s	11 → 21 s	21 → 22 s	22 → 32 s
(a)	10°	10° → 30°	30°	30° → 50°	50°
(b)	50°	50° → 30°	30°	30° → 10°	10°

### 3.5 リアルタイム処理による実験

#### 3.5.1 移動音源による分離性能評価

##### 3.5.1.1 実験条件

2つのラウドスピーカとBSSマイクロホンが配置された実環境で、BSSマイクロホンを用いた移動音源による分離性能評価を行う。部屋の残響時間は200 msである。

評価に用いた2つの音声信号は、音源1は $\theta_1 = -40^\circ$ に固定し、音源2は音源1とは異なる方向から到来する移動音源として、表4に示す2種類の移動パターンで移動させて評価した。評価音の長さは共に32秒であり、2人の男性と2人の女性の合計4人の話者により発声された、研究用ASJ音声コーパス[37]2種類の文章を用いる。これにより、 $(\theta_1, \theta_2)$ それぞれにおいて、12種類の実験ペアを得る。サンプリングレートは8 kHzとする。それぞれの音声と背景雑音の平均SNRは26 dBである。

比較する手法は、(A) 従来のバイナリマスク、(B) 従来のSIMO-ICA (1秒バッファリングデータにより分離行列を学習)、(C) 従来のSIMO-ICA (3秒バッファリングデータにより分離行列を学習)、(D) 提案するBSS (1秒バッファリングデータにより分離行列を学習)、(E) 提案するBSS (3秒バッファリングデータにより分離行列を学習)、の5種類である。提案するBSSにおいては、 $[c_1, c_2, c_3] = [1, 0, 0.5]$ とした。これは、実験環境の背景雑音下では、最も適した(NRRが高いが、CDは低い)パラメータである。

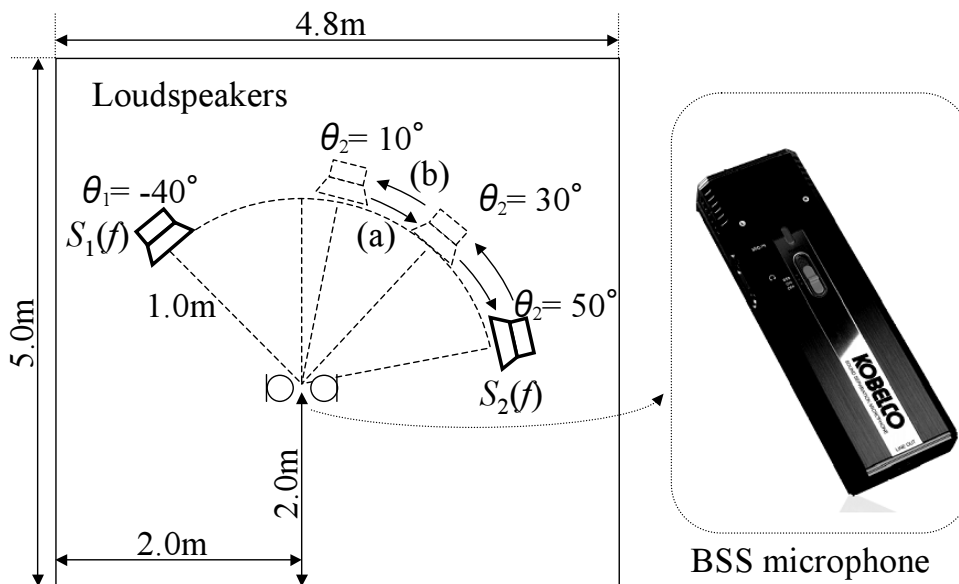


図 16 実験で用いた残響室のレイアウト

### 3.5.1.2 実験結果

図 17 に、時間軸を 0.5 秒刻みに分割して計算された 12 種類の実験ペアの平均セグメンタル NRR を示す。ここでは、セグメンタル NRR の計算に使った分離信号は、音源分離マイクロホン SSM-001 の出力を計算機で録音し、その録音した音をオフラインでセグメンタル NRR として評価した。図 17(a) は時計回りの場合の NRR の結果、図 17(b) は反時計回りの場合の NRR の結果を示す。図 17(a) と (b) の両方において、3 秒バッファリングデータを用いた提案法が 3~32 秒の時間において、従来のバイナリマスクや従来の SIMO-ICA と比較して優れていることがわかる。提案法におけるバッファリングデータの長さ (1 秒と 3 秒の) の違いは、1 秒の場合の方が実験開始直後および音源移動直後にわずかに早く性能の改善が見られるが、静止音源は 3 秒の場合と比較して性能が劣っている。これは音声信号の統計的独立性の評価には、1 秒では短すぎることを示していると思われる。これら結果を考慮して、音源分離マイクロホン SSM-001 に実装した 2 段階 BSS アルゴリズムでは、3 秒バッファリングデータを採用する。



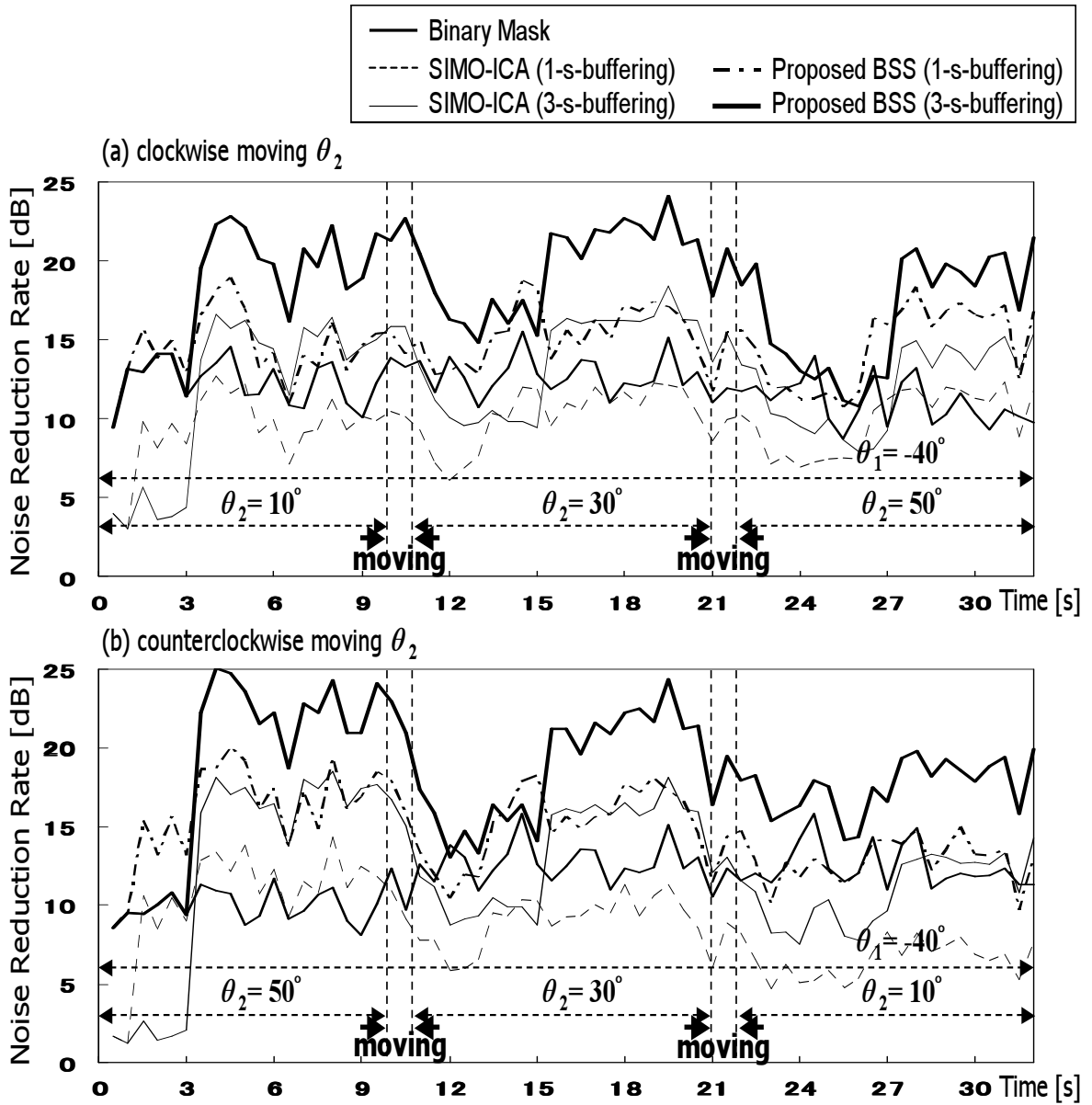


図 17 リアルタイム BSS での実際の録音データを用いて計算した 0.5 秒毎のセグメンタル NRR

## 3.5.2 分離性能のポーラパターン

### 3.5.2.1 実験条件

図 18 に実験条件を示す．2つのラウドスピーカと BSS マイクロホンを実際の部屋に設置して集音した音源信号を用いてリアルタイム音源分離を行う．2つの音源はそれぞれ 30 秒とする．音源 1 は  $\theta_1 = -45^\circ$  に固定し，音源 2 は  $\theta_2 = 0^\circ$  to  $\theta_2 = 360^\circ$  の範囲で  $5^\circ$  刻みに設置して音源分離を行う． $[c_1, c_2, c_3]$  は  $[1, 0, 0.1]$  と  $[1, 0, 0.5]$  の 2 つの条件とし，音源 1 を目的音，音源 2 を雑音として NRR を評価する．

### 3.5.2.2 実験結果

図 19 に実験結果を示す．図を見てわかるとおり，BSS マイクロホンは， $\theta_2 = -5 \sim -185^\circ$  の範囲の右側のエリアに雑音がある場合に高い雑音抑圧性能を発揮することがわかる．一方，BSS マイクロホンは左側のエリアに雑音がある場合には分離を得意としない．これは，(a) ICA 部が音源の角度差が狭い場合に音源分離が難しい，(b) バイナリマスクがうまく機能しない．つまり，左側マイク素子は目的音の音源信号を強く観測し，右側マイク素子は雑音の音源信号を強く観測しなければならないにも関わらず，雑音の音源信号も左側マイク素子側に強く観測されるため，左右マイク素子の観測音量差で機能するバイナリマスクは SIMO-ICA の消し残り雑音抑圧がほとんど機能しないという状態になるのである．この問題は，解決すべき重要な問題であるが，我々の研究グループでは，既にこの問題の解決を狙った改良アルゴリズムの研究も進めている [34] ．

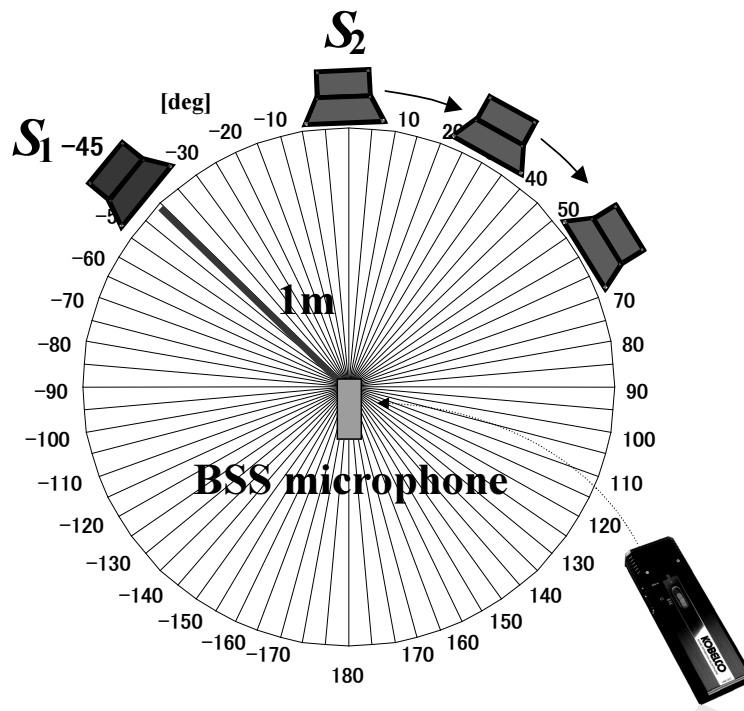


図 18 分離性能のポーラパターン計測条件

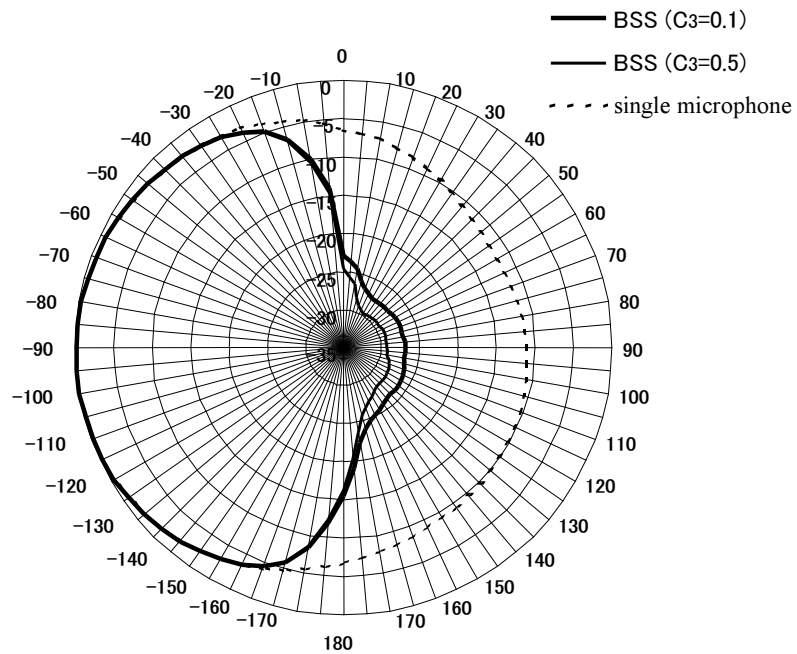


図 19 分離性能のポーラパターン計測結果

### 3.6 本章のまとめ

本章で、高分離性能かつ低歪みな SIMO-ICA と SIMO-BM を結合した 2 段型リアルタイム BSS 手法について述べ、提案法を DSP にリアルタイム実装して実現した音源分離マイクロホン SSM-001 をハードウェア、ソフトウェアの観点から詳細を述べた。また、開発した音源分離マイクロホンの評価として、基本性能評価およびリアルタイム性能評価を行い、開発した音源分離マイクロホンの有効性を示した。

4 章では、さらに音源分離マイクロホンの実用性を高めるための音源抽出法の提案と、本手法およびマイクロホンの有効性を検証するための音源抽出実験の結果について述べる。

## 4. 並列ICAによる雑音推定に基づくブラインド音源抽出

### 4.1 まえがき

本研究の目的は、機械装置の音響診断、携帯電話での音声通話、ロボット・カーナビでの音声認識等、多くのアプリケーションで「聞きたい音」だけを瞬時に取り出すために実用的な装置を開発することであった。その目的から、前章では、SIMO-ICA と SIMO-BM を結合した 2 段型リアルタイム BSS 手法を考案し、音源分離マイクロホン SSM-001 の評価を通じてその有効性を示すところまで行った [38, 39]。

確かに、前述の音源分離マイクロホンは、音声会話等のように目的音と雑音が点音源とみなされる環境では非常に効果的であり、その研究成果はある程度限定的な音響環境では有効であろう。しかし、多くのアプリケーションが想定する実環境では、このような雑音環境は極めて稀である。例えば、携帯電話での音声通話やロボット・カーナビでの音声認識等のシステムにおいては、マイクロホンの近傍に点音源と近似できる目的音声、その周辺に非点音源雑音、つまり、様々な方向から様々な雑音が到来する拡散性雑音が主として存在すると考えられる。よって、これら手法や装置では実環境において十分に性能を発揮できないケースもしばしば存在する。そこで、ここでは、2.2 節で示した混合過程を前提とするのではなく、点音源である目的音一つと、それ以外の雑音は拡散性雑音とする、より実用的な混合過程を対象に考える。つまり、拡散性雑音環境下で目的音をいかに精度よく取り出すかという音源抽出問題を取り扱うことにする。

### 4.2 従来法

#### 4.2.1 ICA に基づく BSS

4.1 節で示した混合過程においても、ICA に基づく BSS 手法は当然有力な手段の一つである。その中でも、特に周波数領域 ICA (Frequency-Domain ICA :

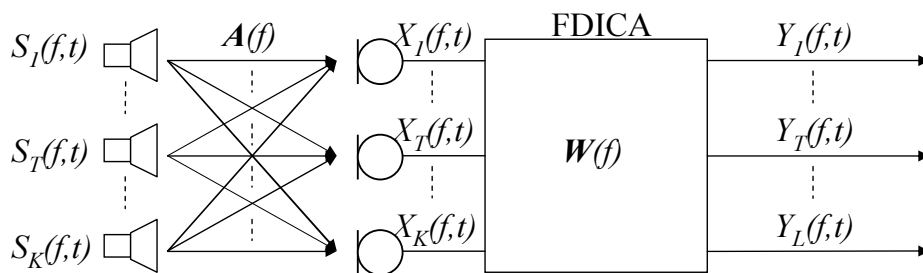


図 20 FDICA に基づく音源分離法

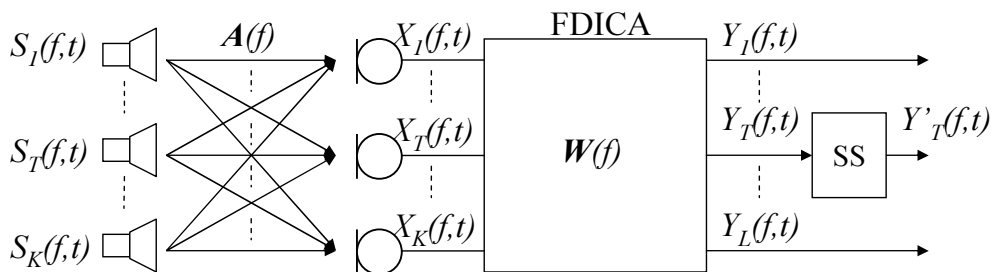


図 21 FDICA と SS の単純接続法

FDICA 以下，FDICA) による BSS 技術が実装上でも有望なアプローチの一つとされている [12, 10, 15, 16] . 本アルゴリズムの詳細は既に 2.3.2 節にて詳述したのでそちらを参照されたい .

既に述べた通り，ICA に基づく BSS 手法は，音声会話のように目的音と雑音が点音源とみなされる環境では非常に効果的である . しかし，実環境では，このような雑音環境は極めて稀であり，実際には様々な方向から様々な雑音が到来する拡散性雑音が主であるために，実環境の多くのケースで性能が大きく劣化する可能性が高い .

#### 4.2.2 ICA に基づく BSS と SS の単純接続

一方、「聞きたい音」だけを抽出する技術として、SS [1] による手法が古くから活用されている。図 21 のように、前述の ICA に基づく BSS 手法で得られた目的音出力  $Y_T(f, t)$  に連結して、 $Y'_T(f, t)$  を得て拡散性雑音に対処することも一つの手法である。この部分の SS は一般に次のように実施される。

$$|Y'_T(f, t)| = \begin{cases} \{|Y_T(f, t)|^2 - \alpha \cdot |N_{Y_T}(f)|^2\}^{\frac{1}{2}} \\ \text{(if } |Y_T(f, t)|^2 - \alpha \cdot |N_{Y_T}(f)|^2 \geq 0) \\ \beta \cdot |Y_T(f, t)| \quad \text{(otherwise),} \end{cases} \quad (34)$$

$$Y'_T(f, t) = |Y'_T(f, t)| \cdot e^{j\arg(Y_T(f, t))} \quad (35)$$

ここで、 $N_{Y_T}(f)$  は目的音が存在しない時間区間で推定した雑音信号、 $Y'_T(f, t)$  は SS 後の出力信号、 $\alpha$  は減算係数、 $\beta$  はフロアリング係数である。これら SS に基づく手法は、定常的な雑音環境下では有効に機能するので、効果的な雑音抑圧手法であると言える。しかし、実環境では、雑音が常に定常であるようなケースもまた極めて稀であり、実環境の多くのケースでは非定常雑音を取り扱うことが要求される。このような場合、SS はその性能を十分に発揮できない。

そこで次節では、雑音が拡散性雑音かつ非定常な雑音であっても機能する BSE 手法を提案する。

### 4.3 アプローチ

拡散性雑音の除去が実用上大きな問題であるが、BSS において有力な手法である ICA に基づく方法では、4.2.1 節でも言及したとおり、拡散性雑音を効果的に除去することは非常に難しい。

これは、実は ICA の動作原理に起因している。図 22 に一例を示すように、ICA による最適化により生成される音源分離フィルタは、点音源に対しては、その音

源方位に深い死角型フィルタを生成することが知られている [41]。一方，拡散性雑音に対しては，様々な方位に対して深い死角を作ることはできず，方向的に広範囲に非常に浅くて緩やかな分離フィルタしか生成されないという現象が起こる。

逆に言えば，4.1 節で前提とした近接点音源と拡散性雑音の環境下において，ICA は目的音声そのものよりも雑音成分の推定に有効であると言える [29]。この現象について，Takahashi ら [42] や大迫ら [43] は，ICA が拡散性雑音に対しては遅延和アレー (Delay-and-sum : DS) を形成し，一方で，目的音声に対しては死角制御型ビームフォーマ (Null Beamforming : NBF) を形成することを理論的に示し，その結果，少ないマイク素子でも高精度な拡散性雑音推定が可能となることに言及している (付録 A にその概要を示す)。

よって，本研究では，この知見をもとに，ICA の使い方に関する着眼点を変える。ICA を使って直接目的音を取るような従来の BSS アプローチを基本とせず，ICA はあくまで「雑音推定器」として考えるのである。この考え方に基づき，拡散性背景雑音に頑健で，DSP 等の専用ハードウェアでリアルタイムに処理可能な実用的な手法を検討する。

#### 4.4 提案法

図 23 に提案法のブロック図を示す。提案法は一つの主マイクロホンと複数の参照マイクロホンから構成されており，主マイクロホンと複数のそれぞれの参照マイクロホンの組合せで 2 入力の ICA 処理を行う構成にしている。既に 4.3 で言及した通り，提案法の基本的な考え方は「ICA の分離音声出力をそのまま用いるのではなく，雑音推定器として用いる」ことであり，全体構成のアイデアは「ICA の雑音推定信号をもとの観測信号から取り除く」というものである。さらに，提案法の大きな特徴は，2 入力 ICA を並列に複数個に分散させる手法を導入することである。この構成の着想は以下の理由による。

1. ICA の役割は，あくまでも雑音推定器であるため，分離音声出力と分離雑音出力 (非音声成分の全て) の 2 出力 ICA で十分である。
2. ICA の演算量の削減効果が期待できる。例えば，マイクが  $K$  個ある場合，



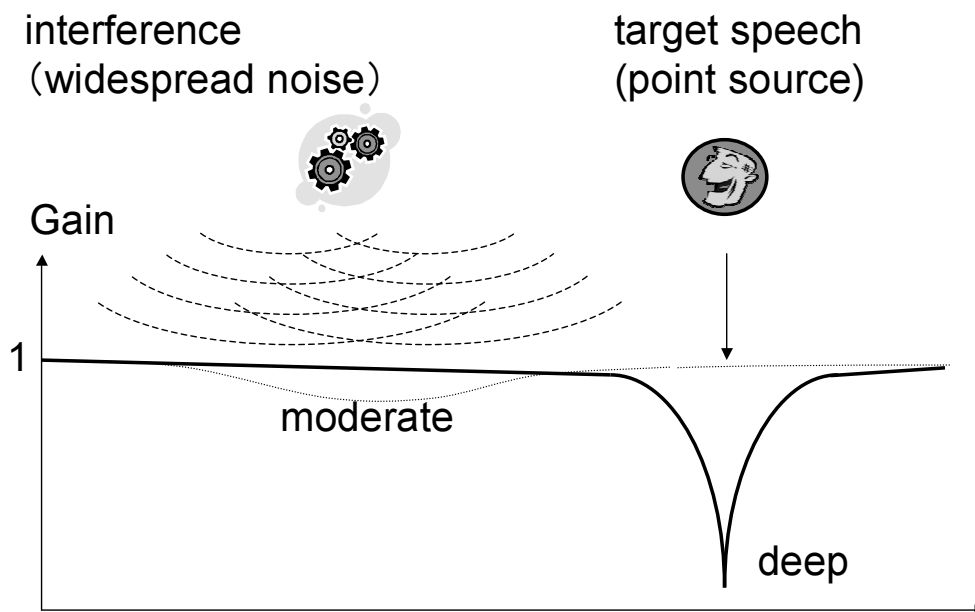


図 22 ICA による死角型フィルタ

普通の  $K$  入力  $K$  出力 ICA を行うと、その行列のサイズは  $K \times K$  である。一方、提案法では、2 入力 2 出力 ICA を  $(K - 1)$  個用いるので、その行列の要素数は  $2 \times 2 \times (K - 1) = 4K - 4$  である。一般に、 $K$  が 3 以上の時、 $K \times K > 4K - 4$  なので、提案法の方が ICA で更新する行列係数の数が少なくてすむ。

また提案法では、複数 ICA の雑音推定結果の統合を、複素振幅の二乗値の重み付け加算にて行う。この際、位相の回復は無視されるが、これは音声処理において位相の復元にあまり重きがおかれないことに由来する。このように、複数 ICA で得られた推定雑音成分のパワー同士の加算で総合的な雑音パワーの推定を行うことにより（位相も考慮して）波形自体の回復を行う場合に比較して雑音抑圧処理の自由度が増し、大幅な雑音低減が可能となる。

さらに、実用性を向上させるために、それぞれのマイクロホンは指向性素子を採用し、参照マイクロホンは主マイクロホンとは異なる方位に指向性を形成して取り囲むようにバランス良く配置させている。これにより、様々な方向から到来

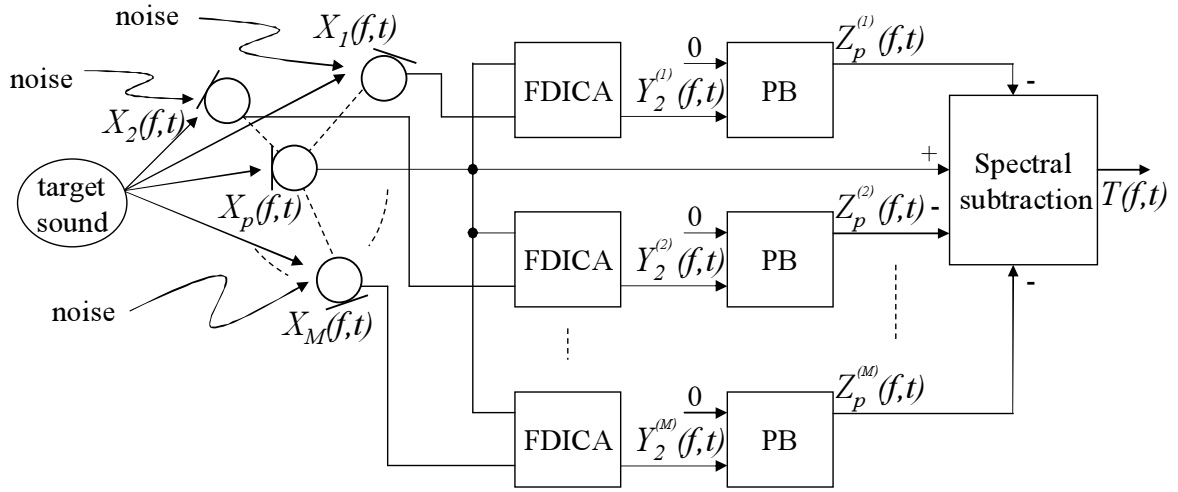


図 23 提案法のブロック図

する分離雑音出力を空間的に網羅して，主マイクロホン方位に概位置する音源抽出性能の頑健性を向上させている．

以下，提案法のアルゴリズムについて，処理の流れにそって説明する．周波数領域の観測信号ベクトルは以下の式で定義される．

$$\mathbf{X}_m(f, t) = [X_p(f, t), X_m(f, t)]^T \quad (36)$$

ここで， $\mathbf{X}_m(f, t)$  は観測信号， $m (= 0, 1, 2, \dots, M)$  は参照マイクロホンインデックス， $X_p(f, t)$  は主マイクロホンの観測信号， $X_m(f, t)$  は参照マイクロホン  $m$  の観測信号である．この  $\mathbf{X}_m(f, t)$  を ICA の入力として拡散性雑音推定を行う．ここでの処理は FDICA を使う．ここでは，次式で与えられる周波数領域の分離行列によって音源分離が行われる．

$$\mathbf{W}_m(f) = \begin{bmatrix} W_{11}^{(m)}(f) & W_{12}^{(m)}(f) \\ W_{21}^{(m)}(f) & W_{22}^{(m)}(f) \end{bmatrix} \quad (37)$$

分離出力  $\mathbf{Y}_m(f, t) = [Y_1^{(m)}(f, t), Y_2^{(m)}(f, t)]^T$  の各要素は互いに独立になるように処理される．この手順は以下のように与えられる．

$$\mathbf{Y}_m(f, t) = \mathbf{W}_m(f) \mathbf{X}_m(f, t) \quad (38)$$

この時， $W_m(f)$  の最適化は FDICA の一般的な以下の反復学習式で行われる．

$$\begin{aligned} W_m^{[i+1]}(f) &= \eta [I - \langle \Phi(Y_m(f, t)) Y_m^H(f, t) \rangle_t] W_m^{[i]}(f) \\ &\quad + W_m^{[i]}(f) \end{aligned} \quad (39)$$

ここで， $I$  は単位行列， $\langle \cdot \rangle_t$  はフレーム平均演算子， $[i]$  は  $i$  番目の反復における値， $\eta$  は更新係数であり， $\Phi(\cdot)$  は適当な非線形関数である．それぞれの FDICA ブロックでは，推定雑音要素のみを残す．つまり，推定目的音信号（ただし実際には精度よく推定されていない） $Y_1^{(m)}(f, t)$  は出力信号  $Y_m(f, t)$  から取り除く．一般的に ICA では，出力のどれが推定雑音要素かは不明だが，ここでは，4.5 節で説明するハードウェア構成の特徴，つまり，主マイクロホンの指向性が常に概目的音方向に向いている点を利用する．具体的には，1) 主マイクロホンの指向性の方位に死角を設定した ICA 初期値を使って ICA 学習時の目的音に対する NBF 要素の順番を合わせる，2) 主マイクロホンと副マイクロホンの指向方位の違いに起因する目的音と雑音の音量差を利用したパーミュテーション解法を利用する，の 2 点の対応により，安定して推定雑音要素を取り出す．この処理は以下のように実施される．

$$U_m(f, t) = [0, Y_2^{(m)}(f, t)]^T \quad (40)$$

次に，PB 法 [44] を用いて推定雑音信号レベルの不確定性を取り除く．その処理は以下のように与えられる．

$$Z_m(f, t) = M_m^+(f) U_m(f, t) \quad (41)$$

ここで， $Z_m(f, t) = [Z_p^{(m)}(f, t), Z_r^{(m)}(f, t)]^T$  は  $m$  番目の推定雑音ベクトルであり，その要素である  $Z_p^{(m)}(f, t)$  と  $Z_r^{(m)}(f, t)$  はそれぞれ主マイクロホンに入力される  $m$  番目推定雑音および参照マイクロホンに入力される推定雑音である．また， $M^+(f)$  は  $W_m(f)$  の Moore-Penrose 型の擬似逆行列である．最後に，音源抽出は SS に

より次のように実施される．

$$|T(f, t)| = \begin{cases} \left\{ |X_p(f, t)|^2 - \sum_{m=1}^M \alpha_m \cdot |Z_p^{(m)}(f, t)|^2 \right\}^{\frac{1}{2}} \\ \text{(if } |X_p(f, t)|^2 - \sum_{m=1}^M \alpha_m \cdot |Z_p^{(m)}(f, t)|^2 \geq 0) \\ \beta \cdot |X_p(f, t)| \quad \text{(otherwise),} \end{cases} \quad (42)$$

$$T(f, t) = |T(f, t)| \cdot e^{j \arg(X_p(f, t))} \quad (43)$$

ここで， $T(f, t)$  は提案法の最終出力， $\alpha_m$  は減算係数， $\beta$  はフロアリング係数である．例えば， $(\sum_{m=1}^M \alpha_m) \simeq 1$ ， $\beta \ll 1$  とすることで，効果的な雑音除去が可能になる．後述のマイクロホンでは， $M = 2$  とし， $\alpha_m = 0.5$ ， $\beta = 0.1$  としている．なお，ここで重要な点は  $Z_m(f, t)$  が時間フレーム  $t$  の関数になっていることである．つまり，従来の SS のような定常的かつ代表的な推定雑音を用いていない．それゆえに提案法は非定常的な雑音も取り扱うことが可能である．

## 4.5 ブラインド音源抽出マイクロホン

### 4.5.1 概要

BSE マイクロホンの外観は図 8 に示した BSS マイクロホンと同様である．また，内部基板を図 24 に，さらに，ハードウェアブロック図を図 25 に示す．また，本マイクロホンの概要を表 5 にまとめた．図 24 からわかるように，主マイクロホン素子 1 個と参照マイクロホン素子 2 個を使用しており，提案法において  $M = 2$  の場合を評価できる．マイクロホン素子は全てスター精密製の EBA-06G-L を使用しており，単一指向性素子である．その指向方向は，主マイクロホンを  $0^\circ$  とした場合，左側の副マイクロホンが  $+90^\circ$  方向，右側の副マイクロホンが  $-90^\circ$  方向に設定されている．主マイクロホンは，常に目的音の方向に概向していると仮定している．これらマイクロホンから入力された入力信号は基板上の DSP にて信号

処理され、ライン出力とヘッドホン出力に同時に出力される。マイコンはスイッチ類のコントロールのみに使用しており、フラッシュROMはDSPプログラムを格納し、SD-RAMは信号処理演算のワークメモリの一部として利用されている。電源は単3乾電池2本であり、本アルゴリズムの動作で5時間程度の連続使用が可能である。

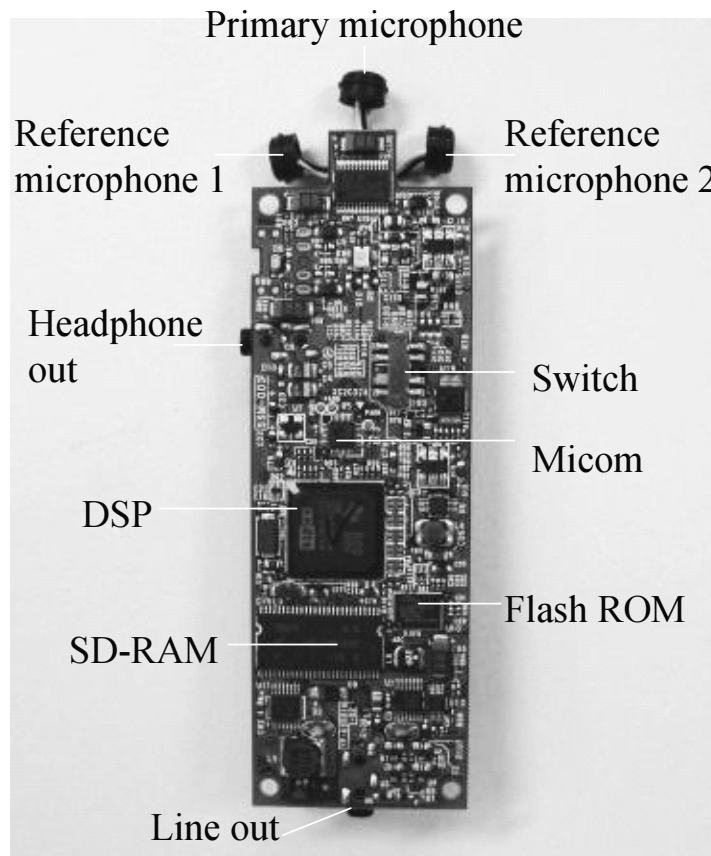


図 24 音源抽出マイクロホンの内部基板

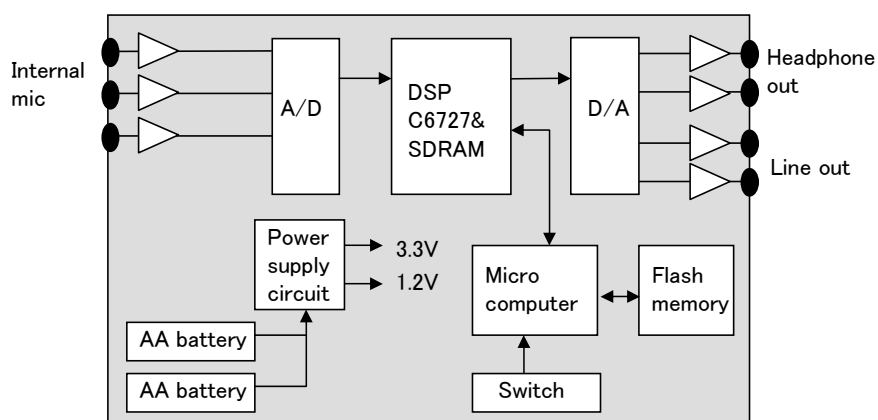


図 25 音源抽出マイクロホンのハードウェアブロック図

表 5 音源抽出マイクロホン概略仕様

プロセッサ	TI 社製 DSP TMS320VC6727 (動作周波数 300 MHz)
入力	主マイクロホン素子 (1 個) 参照マイクロホン素子 (2 個)
出力	ライン出力, ヘッドホン出力
サンプリング周波数	16 kHz
電源	単 3 乾電池 × 2
搭載メモリ	Flash ROM: 8 MB (約 430 KB 使用)
	SDRAM: 128 MB (約 3.9 MB 使用)
大きさ	136 mm (H) × 45 mm (W) × 27 mm (D)
重量	125 g (including battery)

#### 4.5.2 リアルタイム実装

図 26 に提案法のリアルタイム実装フローのブロック図を示す．処理手順は以下の通りである．

1. 主マイクロホンおよび参照マイクロホンへの入力信号を 2048 サンプル（フレーム）毎の FFT 処理により時間-周波数信号へ変換する．
2. ある一定時間長（図 26 中の  $T_s$ ）の FFT 変換信号が揃った段階で参照マイクロホン毎に FDICA による  $W_m(f)$  の最適化処理を行い，計算できた段階（図 26 中の  $T_l$  後）で即座に分離フィルタ係数の更新を行う．
3. 分離フィルタによるフィルタリング処理後の信号に PB 処理を行う．
4. PB 処理後の信号を重み付け加算した信号を用いて SS 処理を施す．
5. SS 処理後の信号を逆 FFT 処理とオーバーラップ加算処理により，時間信号に戻す．

図 26 にも示す通り，FDICA による分離フィルタの最適化処理は瞬時に処理できず， $T_s + T_l$  時間の遅延が生じる．一般に，ICA による分離フィルタの最適化処理はリアルタイムシステムを実現する上で問題となるが，フィルタ最適化処理遅延の間は過去時間で計算された分離フィルタ係数でフィルタリングされるため，PB 処理とスペクトル減算とともに遅延なく現時間のフレーム信号に適用されて目的音抽出は瞬時に行われる．なお，本マイクロホンでは，(39) 式に示される更新式により，100 回の繰り返し学習が行われ，搭載の DSP を用いることにより，図 26 中の  $T_s$  を 3 秒として，図 26 中の  $T_l$  を約 0.5 秒で計算できる．よって，FDICA のフィルタ更新周期は 3 秒，フィルタ更新遅延は約 3.5 秒である．

また，本マイクロホンでは，対象とする目的音源が音声のように断続的な場合も想定している．つまり，主マイクロホンに含まれる信号パワー変動を観測して  $T_s$  の 3 秒での目的信号の有無を判定し，目的音がない場合には，分離フィルタ更新をリセット（初期値へ戻す）することにより，性能の大幅な劣化を防止している点も追記しておく．

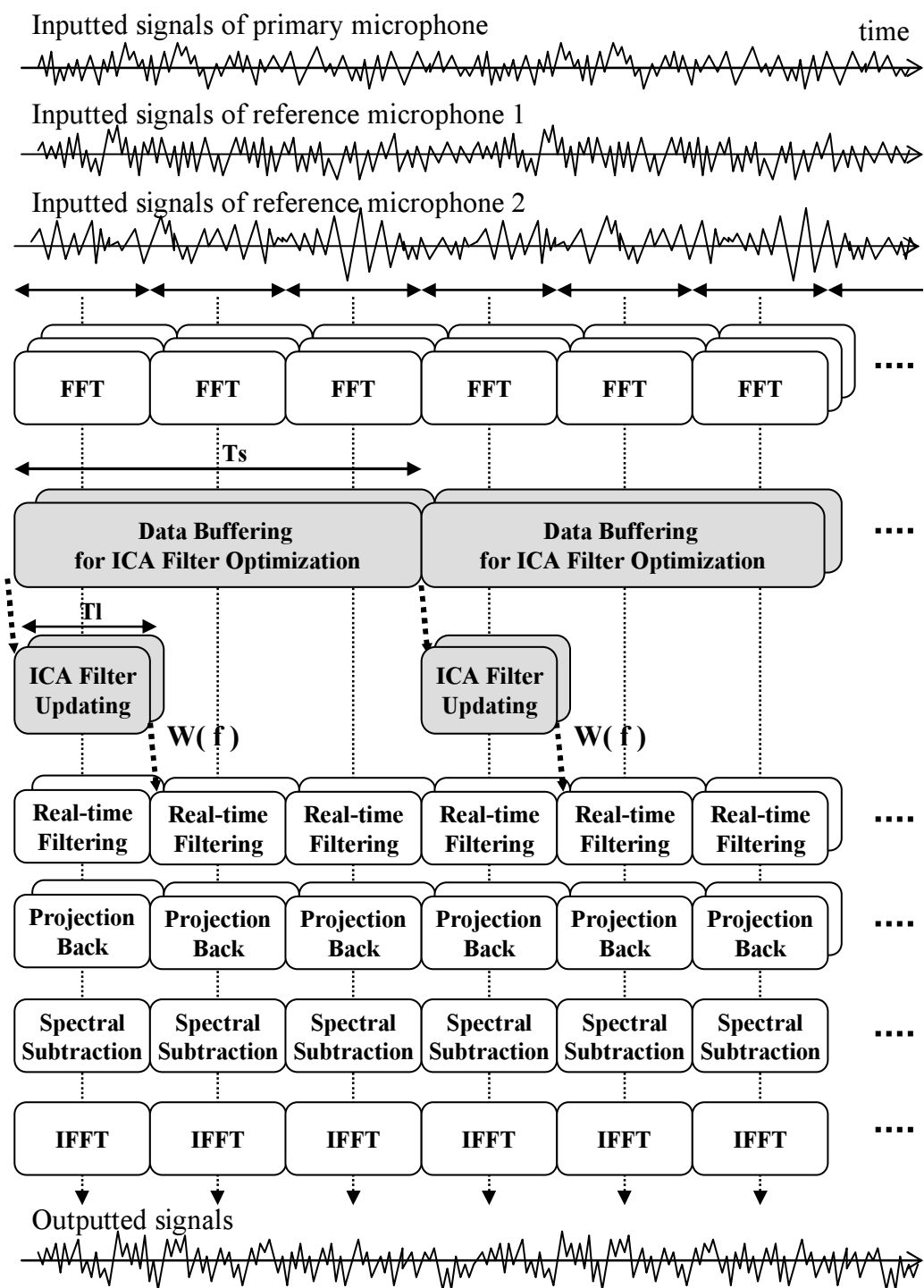


図 26 提案法のリアルタイム処理フロー



表 6 提案法リアルタイム実装時のリソース

ROM 量	430 kByte
RAM 量	3.88 Mbyte
処理量 (CPU 占有率)	40 % (120/300 MHz) @ 16 kHz フィルタ更新処理無の場合 (有の場合は 100 %)

#### 4.5.3 リソース

前節で説明したリアルタイム実装において、各リソースを表 6 にまとめる。C 言語によるソースプログラムを開発し、DSP 開発の統合環境である CCS ver.3 で標準モードでのコンパイルを行う。処理中最も負荷が高くなるケースは、バックグラウンドで動作するフィルタ更新処理が実施される場合であり、この時 MPU の占有率はほぼ 100%に達する。

## 4.6 音源抽出実験

### 4.6.1 実験条件

ここでは、開発したマイクロホンに実装された提案法の音源抽出性能の評価について述べる。評価指標は、(a) 出力 SNR-入力 SNR として定義される雑音抑圧量 (NRR) [16]、(b) 目標信号と抽出信号のスペクトル包絡距離を評価するケプストラム歪 (CD)、(c) NRR と CD 双方に関係する主観品質相当の音質評価値 (Perceptual Evaluation of Speech Quality Most Opinion Score - Listening-only Quality for Objective : PESQ MOS-LQO) で評価する。評価には 16 kHz でサンプリングされた実際の録音信号を用いた。まず、実験のベースとなる環境として図 27 のように、目的信号としてマイクロホン正面 ( $\theta_1 = 0^\circ$  に固定) 1 m の距離から話された女性ユーザ音声  $S_1$  を用い、妨害音声として周囲  $360^\circ$  に  $5^\circ$  刻みで順番に配置された男性ユーザ音源  $S_2$  を用いた。また、拡散性雑音として次の 3 種類を集音した。(I) 白色雑音、(II) 音楽雑音、(III) 駅雑音である。拡散性雑音は (I) 白色雑音と (II) 音楽雑音に関しては、図 28 のようにマイクロホン周囲に配置された 24 個のラウドスピーカを配置して、それぞれのラウドスピーカからそれぞれ相関のない雑音を発生させて集音した。これら信号は Audio/Acoustics Technical CD に集音された白色雑音、男女アナウンス音声および音楽 (NEW WORLD SYMPHONY) で 30 秒分を用いた。この集音雑音を図 27 に示すベースとなる環境に雑音レベルを調整の上で加算した。(III) 駅雑音に関しては、図 29 に示す駅環境で実際の雑音の集音を行った。雑音は電車をはじめ、券売機、改札、車、会話音声、足音、など様々な雑音を含んでおり、非定常かつ非点音源である。大きさはおよそ  $62 \pm 5$  dBA 程度であった。この集音雑音も図 27 のベースとなる環境に雑音レベルを調整の上で加算した。

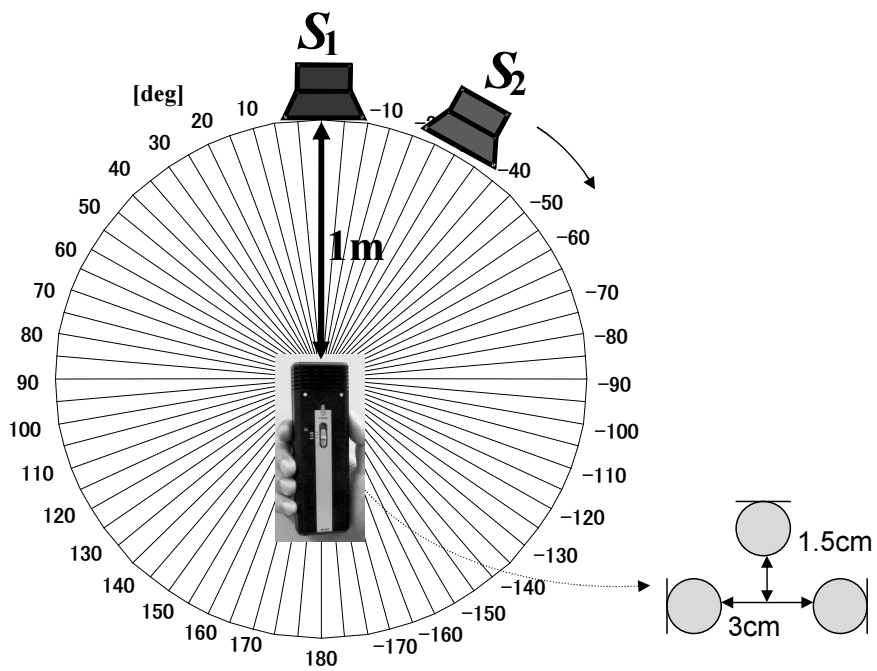


図 27 目的音声と妨害音声の集音条件

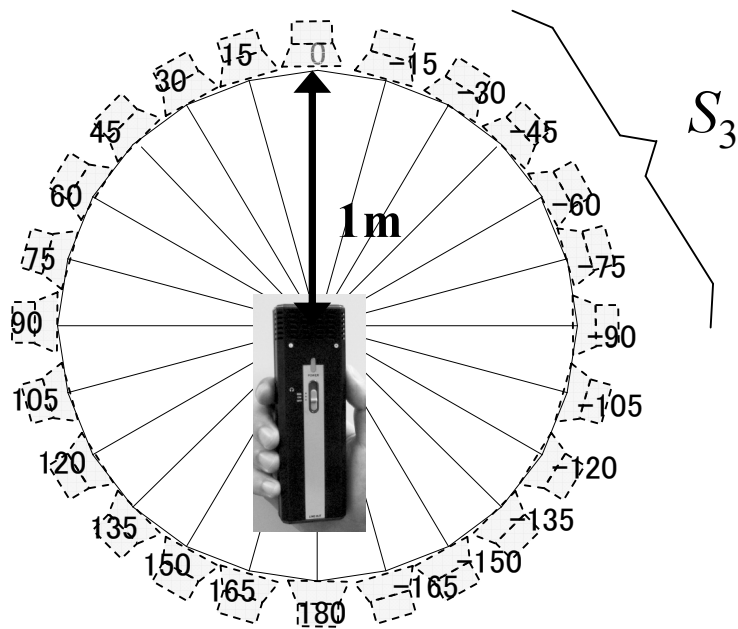
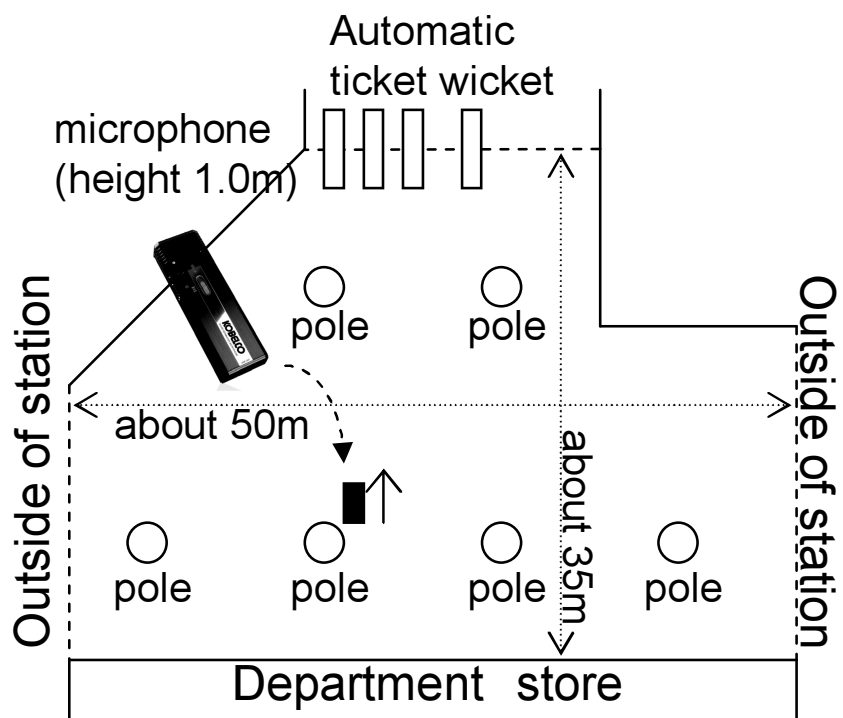


図 28 拡散性雑音の集音条件



(a) Recording environment



(b) Scenery from the recording point

図 29 環境雑音を集音した駅環境

#### 4.6.2 実験結果

実験は3通りの雑音レベルの加算条件で行った。 $S_1$ と駅雑音の平均SNRをそれぞれ(1)20 dB,(2)10 dB,および(3)0 dBである。また,実験では以下の4手法を比較した:(A)主マイクロホン音から2つの参照マイクロホン音を単純にSSした3chSS,(B)全てのマイクロホン入力を用いた従来3chFDICA(図20参照),(C)従来3chFDICAの目的音出力に単純なSSを接続した3chFDICA+1chSS(図21参照),(D)提案法,である。なお,(C)の1chSSは,目的音出力中の雑音区間1秒分を選別してスペクトル平均値を計算して用いた。実験の結果,図30,図33,図36のそれぞれに示す(1)20 dBの条件下において,提案法はNRRで全ての手法よりもやや優れる一方,CDではICAを含む手法にやや劣る。PESQ MOS-LQOではほぼ同等で,総合的優劣はICAを含む手法と同等程度といえる。次に図31,図34,図37に示す(2)10 dBの条件下では,提案法は全ての手法にNRRでさらに大きく勝っている。CDでやや劣っているものの,PESQ MOS-LQOでも大きく優れている。さらに図32,図35,図38に示す(3)0 dBの条件下では,さらにその優位性が際立って現れている。この結果より,提案法はBSE機能において,白色雑音,音楽雑音,駅雑音等の性質の異なる様々な拡散性雑音環境下でも頑健に機能し,他の従来手法と比べて優れた性能を発揮することが確認できた。

以上,客観評価指標を用いて提案システムを評価したが,著者らの主観的な判断においても,ほぼPESQ値に沿う傾向が確認されていることを言及しておく。つまり,提案法は,若干の歪みを生じつつも非常に大きな雑音抑圧量を達成しており,総合的には好ましい雑音抑圧結果が得られていると感じられた。なお,精密な主観評価に関しては,限られた開発期間の制約により実施することが困難であったが,今後の課題であるといえる。

Target speech  $S_1$  :  $0^\circ$   
 Interference  $S_2$  :  $-180^\circ \sim 180^\circ$   
 Widespread noise  $S_3$  : White noise (SNR 20 dB)

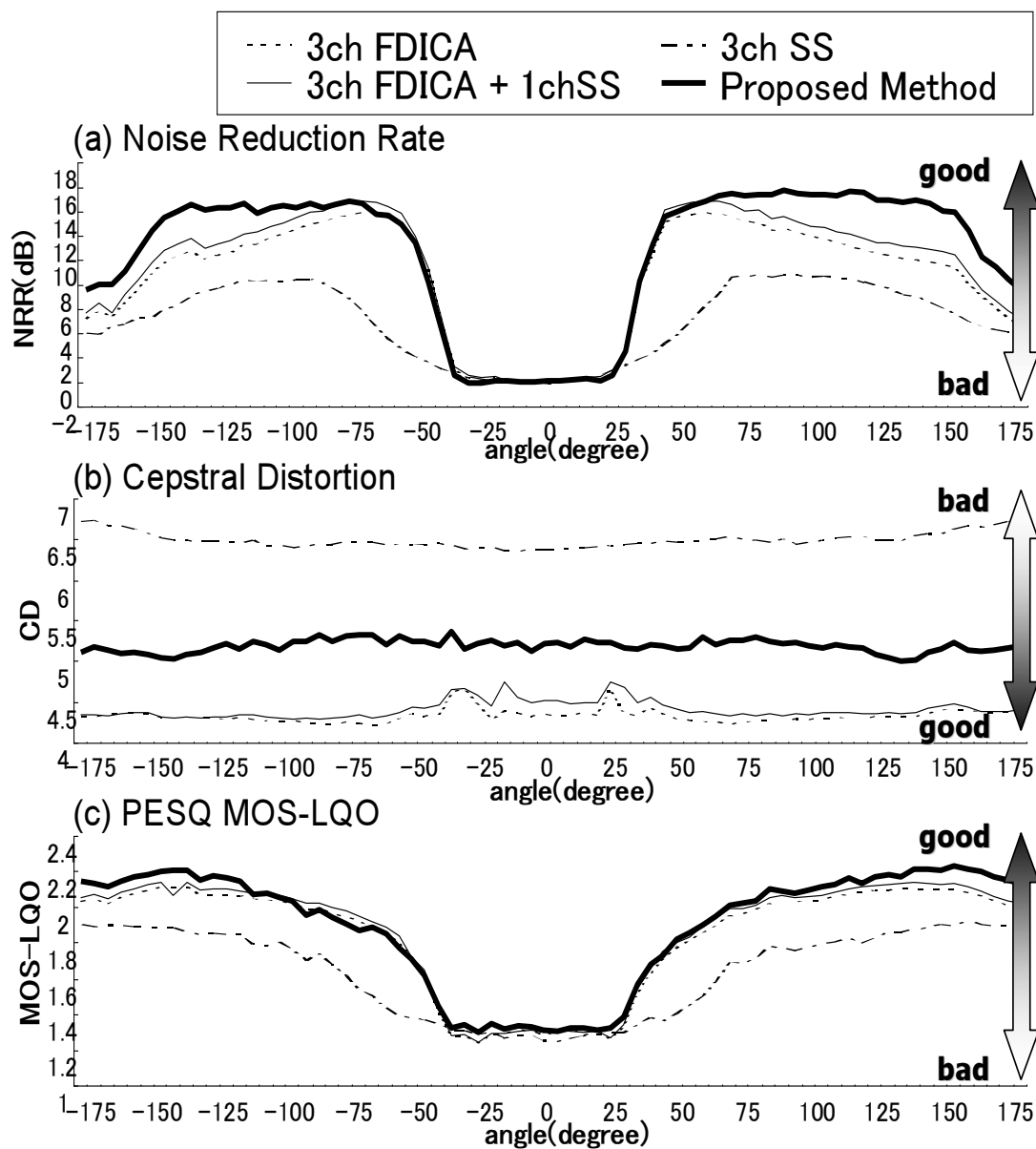


図 30 拡散性雑音を白色雑音 SNR 20 dB とした場合の音源抽出実験結果:(a) 雑音抑圧量, (b) ケプストラム歪, (c) PESQ MOS-LQO . 横軸は妨害音声  $S_2$  の方位を表す .

Target speech  $S_1$  :  $0^\circ$   
 Interference  $S_2$  :  $-180^\circ \sim 180^\circ$   
 Widespread noise  $S_3$  : White noise (SNR 10 dB)

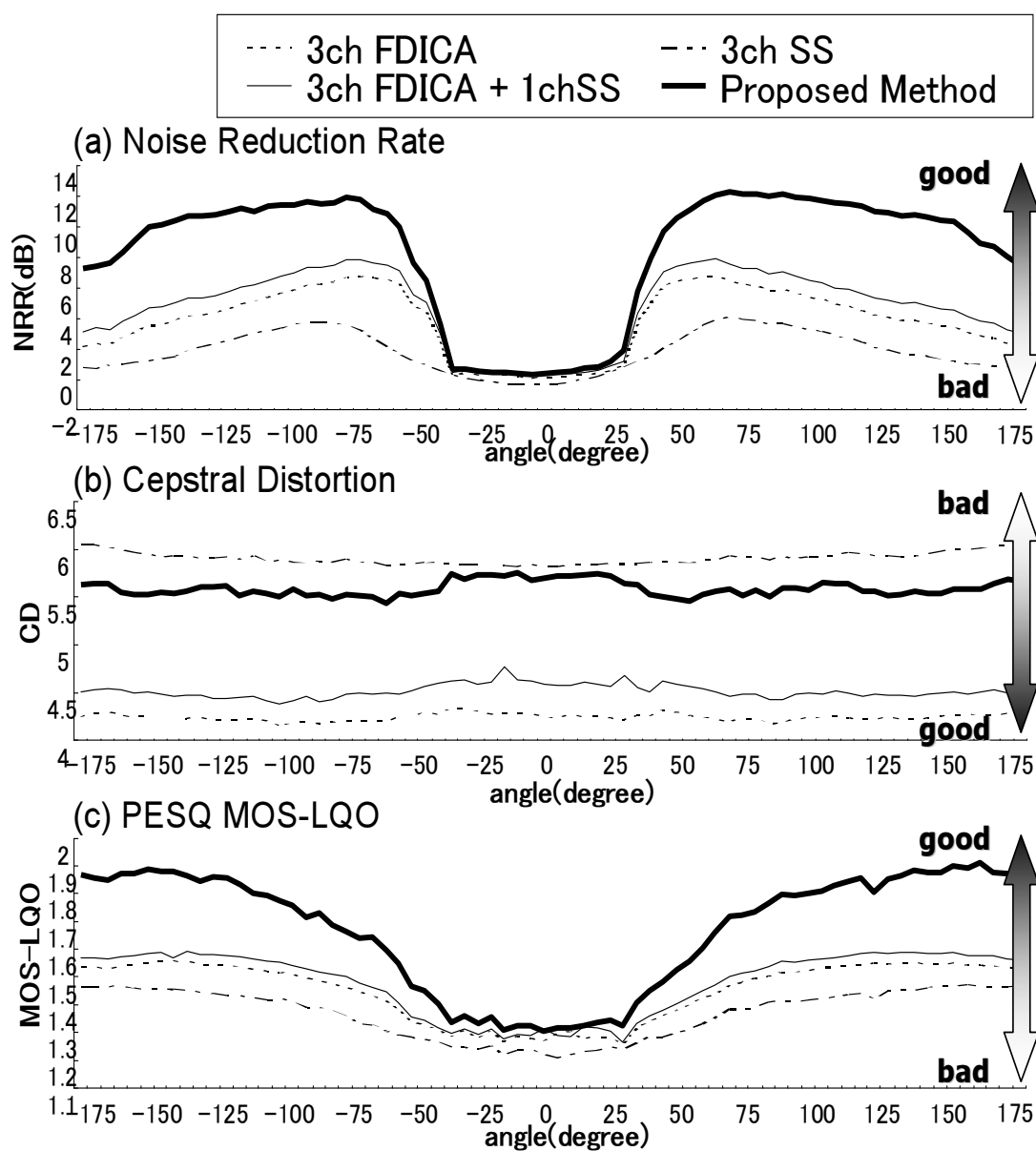


図 31 拡散性雑音を白色雑音 SNR 10 dB とした場合の音源抽出実験結果:(a) 雑音抑圧量, (b) ケプストラム歪, (c) PESQ MOS-LQO . 横軸は妨害音声  $S_2$  の方位を表す .

Target speech  $S_1$  :  $0^\circ$   
 Interference  $S_2$  :  $-180^\circ \sim 180^\circ$   
 Widespread noise  $S_3$  : White noise (SNR 0 dB)

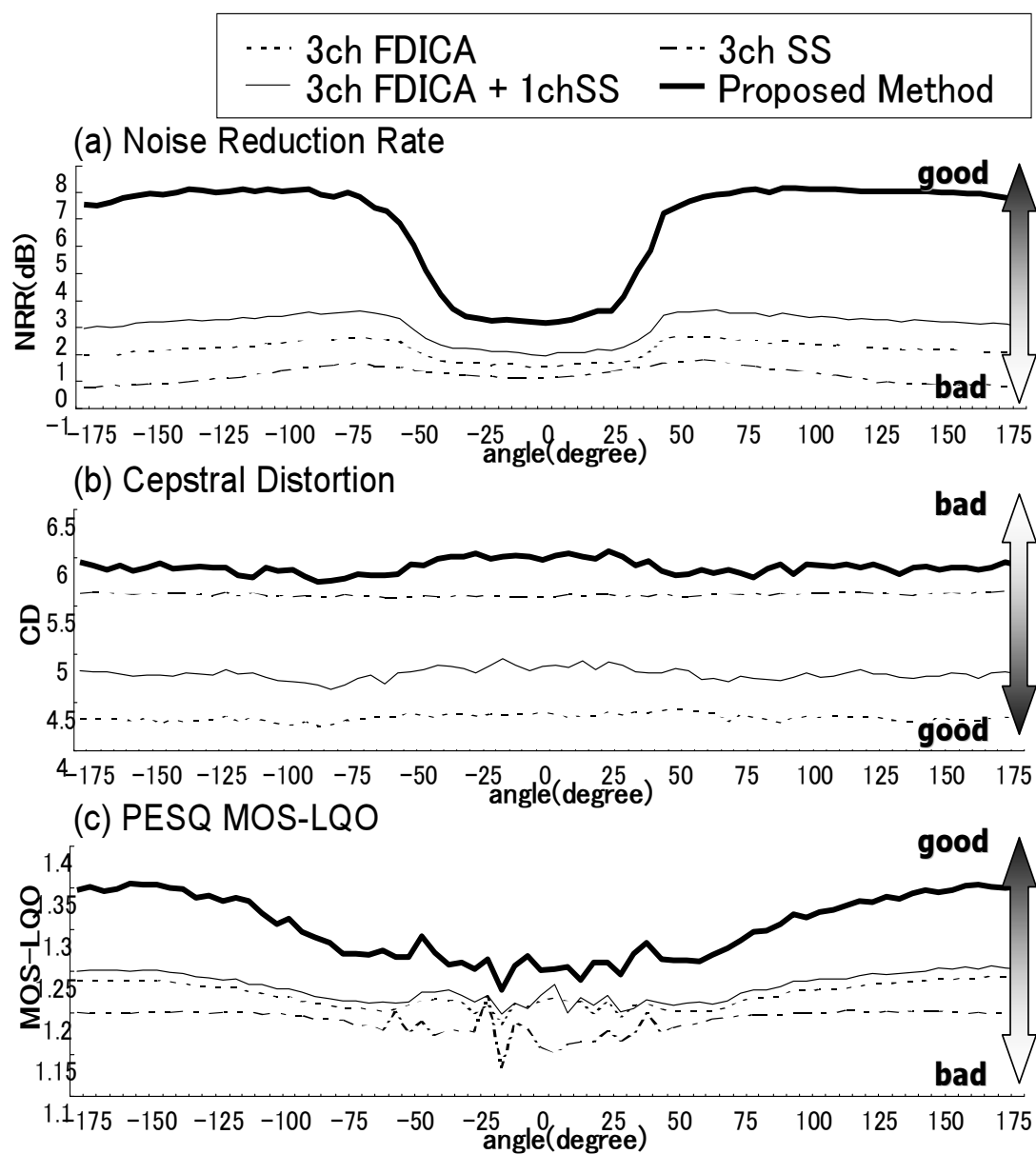


図 32 拡散性雑音を白色雑音 SNR 0 dB とした場合の音源抽出実験結果:(a) 雑音抑圧量, (b) ケプストラム歪, (c) PESQ MOS-LQO . 横軸は妨害音声  $S_2$  の方位を表す .



Target speech  $S_1$  :  $0^\circ$   
 Interference  $S_2$  :  $-180^\circ \sim 180^\circ$   
 Widespread noise  $S_3$  : Music noise (SNR 20 dB)

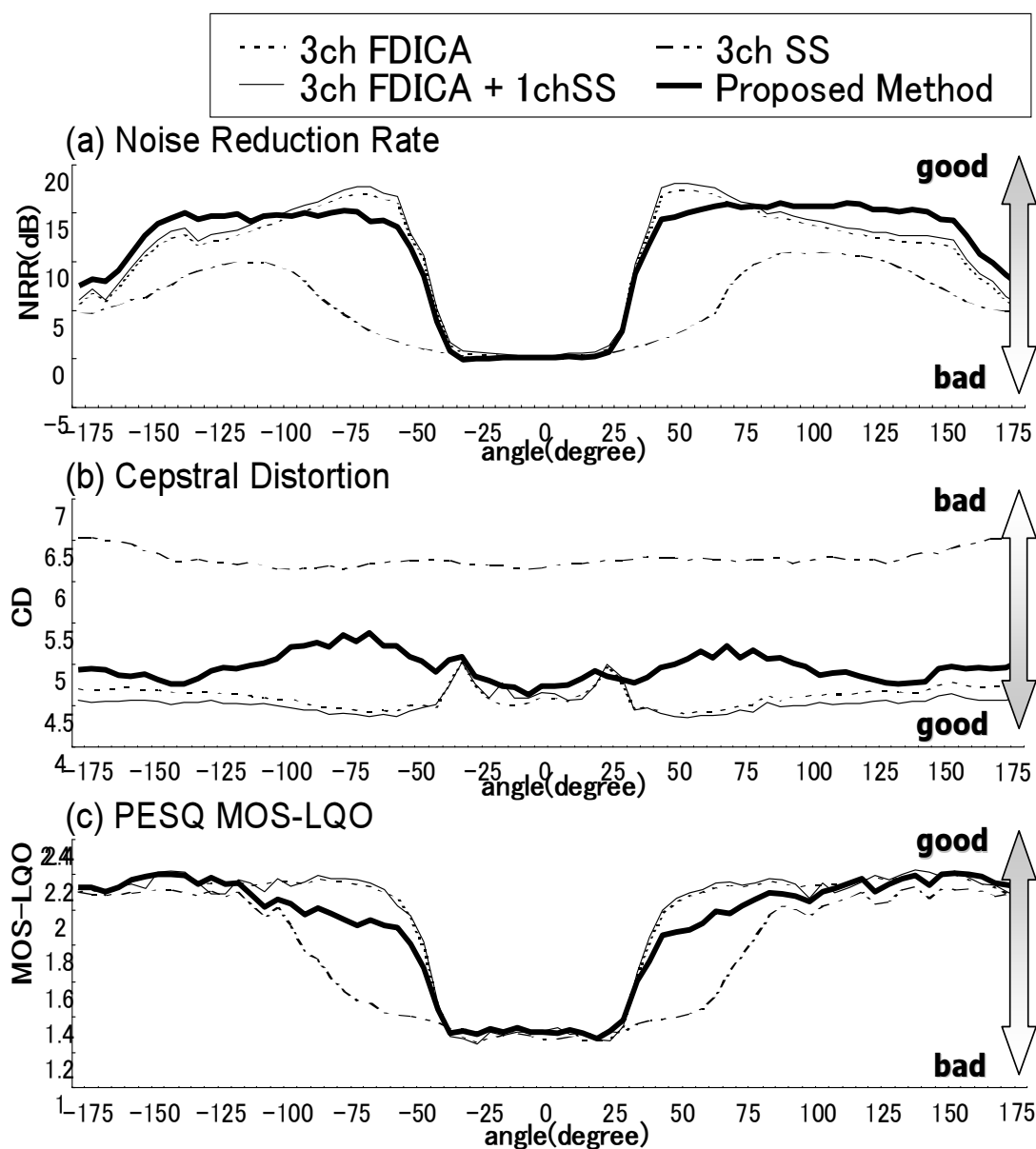


図 33 拡散性雑音を音楽雑音 SNR 20 dB とした場合の音源抽出実験結果:(a) 雑音抑圧量, (b) ケプストラム歪, (c) PESQ MOS-LQO . 横軸は妨害音声  $S_2$  の方位を表す .

Target speech  $S_1$  :  $0^\circ$   
 Interference  $S_2$  :  $-180^\circ \sim 180^\circ$   
 Widespread noise  $S_3$  : Music noise (SNR 10 dB)

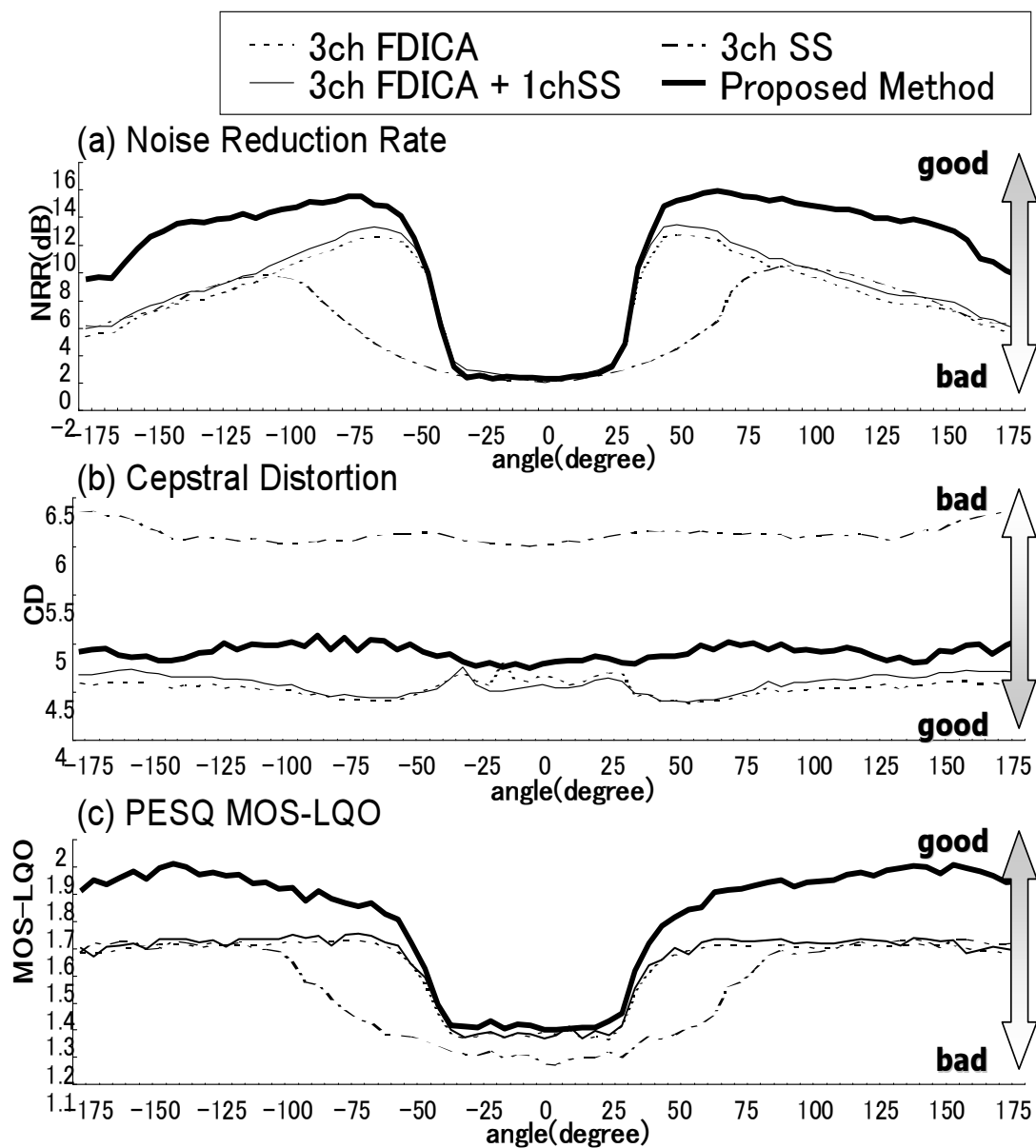


図 34 拡散性雑音を音楽雑音 SNR 10 dB とした場合の音源抽出実験結果:(a) 雑音抑圧量, (b) ケプストラム歪, (c) PESQ MOS-LQO . 横軸は妨害音声  $S_2$  の方位を表す .

Target speech  $S_1$  :  $0^\circ$   
 Interference  $S_2$  :  $-180^\circ \sim 180^\circ$   
 Widespread noise  $S_3$  : Music noise (SNR 0 dB)

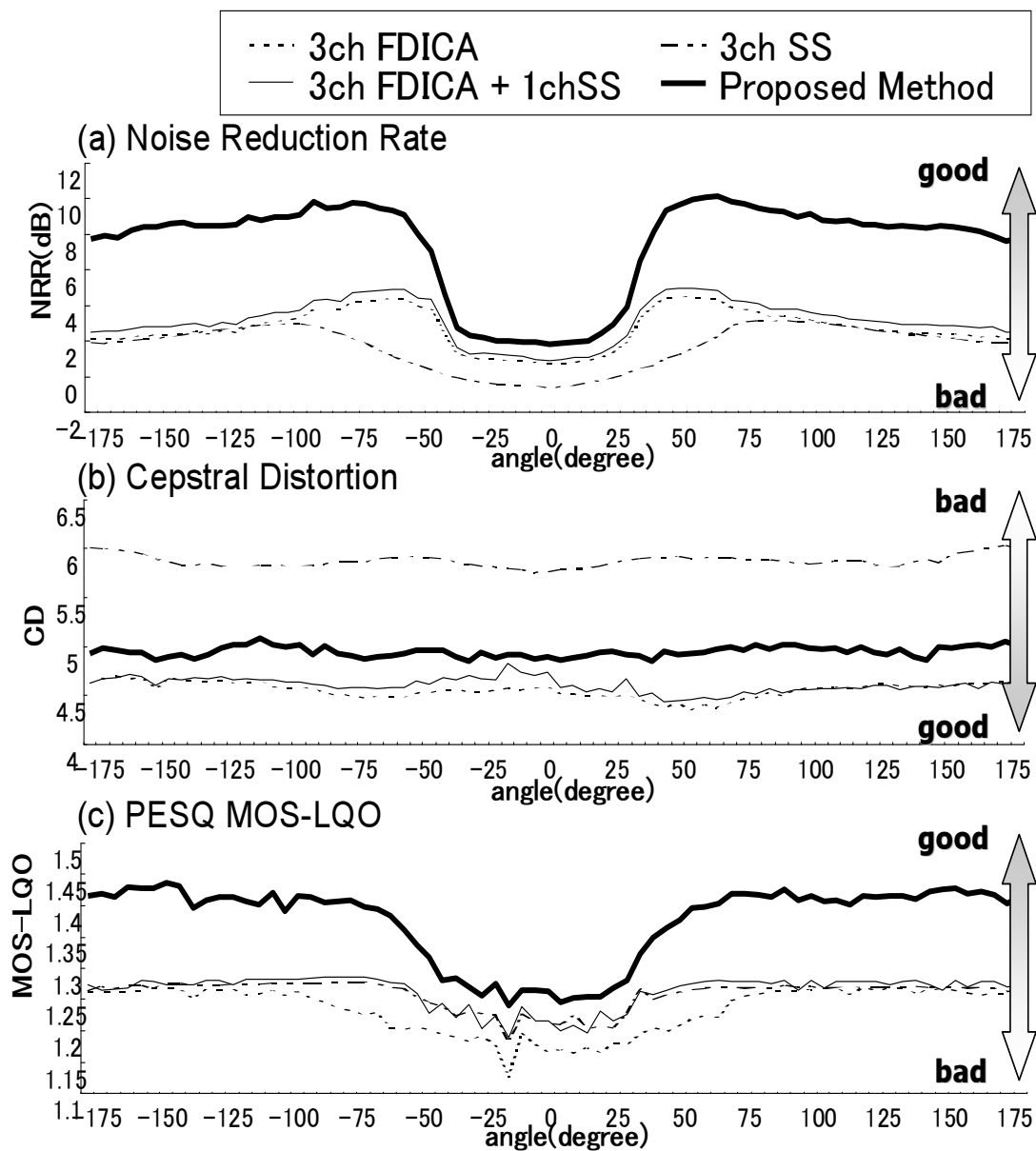


図 35 拡散性雑音を音楽雑音 SNR 0 dB とした場合の音源抽出実験結果:(a) 雑音抑圧量, (b) ケプストラム歪, (c) PESQ MOS-LQO . 横軸は妨害音声  $S_2$  の方位を表す .

Target speech  $S_1$  :  $0^\circ$   
 Interference  $S_2$  :  $-180^\circ \sim 180^\circ$   
 Widespread noise  $S_3$  : Station noise (SNR 20 dB)

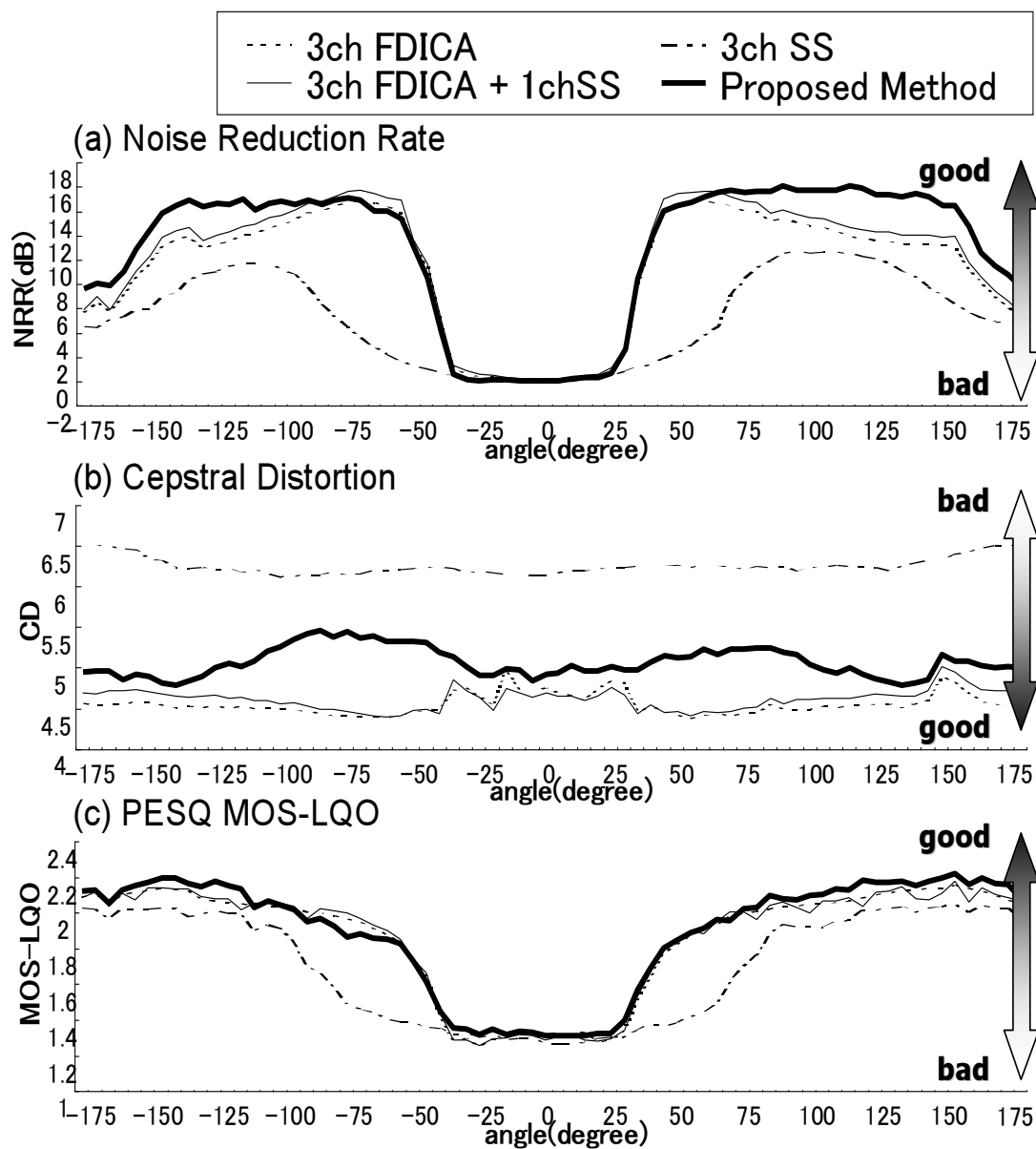


図 36 拡散性雑音を駅雑音 SNR 20 dB とした場合の音源抽出実験結果:(a) 雑音抑圧量, (b) ケプストラム歪, (c) PESQ MOS-LQO . 横軸は妨害音声  $S_2$  の方位を表す .

Target speech  $S_1$  :  $0^\circ$   
 Interference  $S_2$  :  $-180^\circ \sim 180^\circ$   
 Widespread noise  $S_3$  : Station noise (SNR 10 dB)

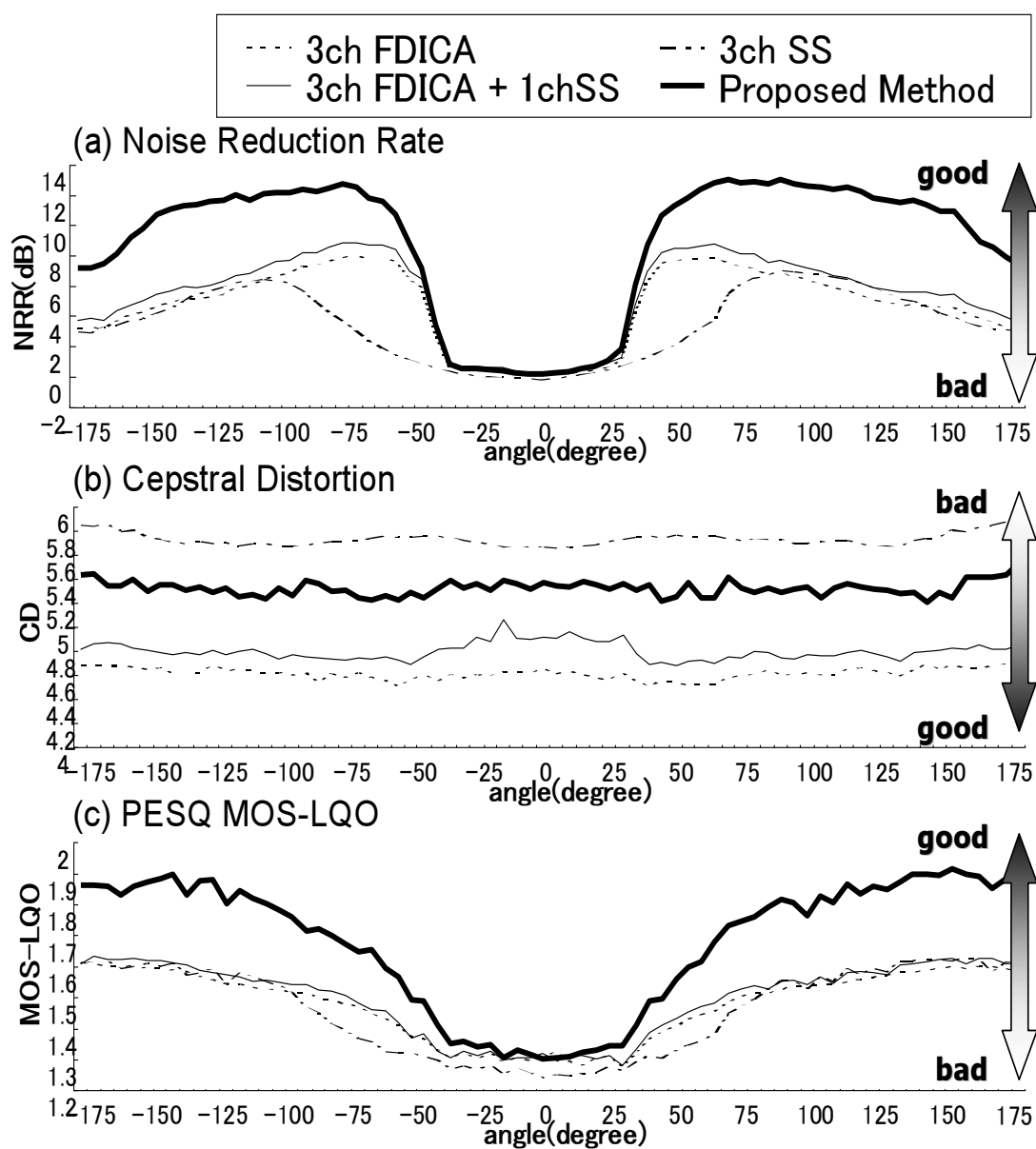


図 37 拡散性雑音を駅雑音 SNR 10 dB とした場合の音源抽出実験結果:(a) 雑音抑圧量, (b) ケプストラム歪, (c) PESQ MOS-LQO . 横軸は妨害音声  $S_2$  の方位を表す .

Target speech  $S_1$  :  $0^\circ$   
 Interference  $S_2$  :  $-180^\circ \sim 180^\circ$   
 Widespread noise  $S_3$  : Station noise (SNR 0 dB)

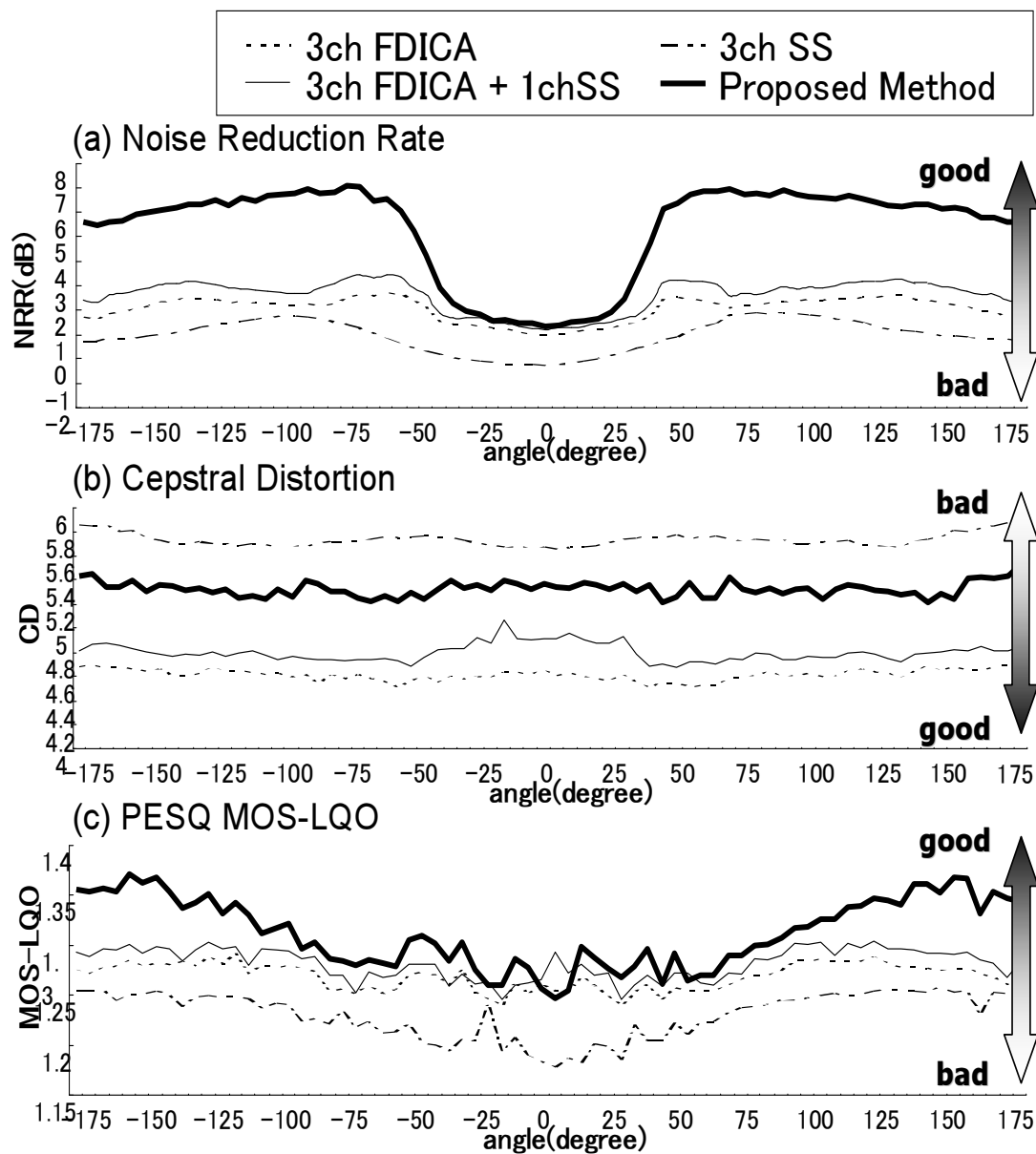


図 38 拡散性雑音を駅雑音 SNR 0 dB とした場合の音源抽出実験結果:(a) 雑音抑圧量, (b) ケプストラム歪, (c) PESQ MOS-LQO . 横軸は妨害音声  $S_2$  の方位を表す .

#### 4.7 本章のまとめ

本章では、並列に動作する複数の ICA と PB および SS に基づいた新しい BSE 手法を提案し、さらに、本提案法を DSP 上に実装して装置化した BSE マイクロホンを用いて、様々な拡散性雑音下で従来手法よりも高い BSE 性能を発揮することを確認した。今後の課題としては、音源抽出性能の向上やケプストラム歪に代表される音質改善等をさらに進めて様々なアプリケーションで要求される実用性能にブラッシュアップしていく必要がある。

## 5. 結言

### 5.1 まとめ

本論文では、実音響環境にて動作する高精度かつ実用的な音源分離マイクロホンの研究開発について論じた。特に、ICA に基づく BSS の高精度化及びリアルタイム化の提案、及び、より実雑音環境に適合した音源分離を実現するため並列に動作する複数 ICA に基づく BSE 手法を提案した。これらのアルゴリズムは、ハードウェア上にてリアルタイム音源分離マイクとして実装され、実機による性能評価が行われた。

第 1 章では、本研究の背景、本研究の目的である「聞きたい音だけを聞くことができる装置の開発」に関する狙いについて述べた。

第 2 章では、まずはじめに、本研究において前提とする混合過程の定式化を行った。また、音源分離の従来法として、バイナリマスクに基づく BSS、ICA に基づく BSS について述べ、さらにこれら手法に関するリアルタイム BSS システムについても述べた。

第 3 章では、高分離性能かつ低歪みで、リアルタイム処理可能な手法として、SIMO-ICA と SIMO-BM を結合した 2 段型リアルタイム BSS 手法について述べ、提案法を DSP にリアルタイム実装して実現した音源分離マイクロホン SSM-001 をハードウェア、ソフトウェアの観点から詳細を述べた。また、開発した音源分離マイクロホンの評価として、基本性能評価およびリアルタイム性能評価を行い、開発した音源分離マイクロホンの有効性を示した。

第 4 章では、さらに実用性の高い手法として、並列に動作する複数の ICA と PB および SS に基づいた新しい BSE 手法を提案し、さらに、本提案法を DSP 上に実装して装置化した BSE マイクロホンをを用いて、様々な拡散性雑音下で従来手法よりも高い BSE 性能を発揮することを確認した。

本研究を通じて開発した、BSS マイクロホン、および BSE マイクロホンは、音源分離・音源抽出性能、ポケットサイズにまで小型化した開発で、新たな音声入力フロントエンドとしての可能性を大いに示すことができたとと言える。



## 5.2 今後の課題

本研究で得られた成果の延長線上には多くのアプリケーションへの応用が期待できる。しかし、実用化へ向けては、まだ多くの技術課題が残されているのも事実である。多くの方々から本技術の将来性を期待され、アプリケーションを想定して多くの評価を頂いて貴重な意見を頂戴した。もちろん、個々のアプリケーション毎にそれぞれ技術課題やレベルは異なるが、主なアプリケーション毎に課題をまとめると、表 7 のようになる。今後は、これら技術課題に向けた研究開発による課題解決が必要となってくると思われる。なお、課題の一つである音声認識性能向上に関しては、BSS マイクロホンと音声認識エンジンを接続した場合の音声認識性能に相当する評価を既に行っており、付録 B に記載しておくので、ご参考願いたい。

表 7 アプリケーション毎の課題

アプリケーション	基本課題	2次課題
音響設備診断	高残響下での分離性能向上 高周波帯域への拡張 ~ 32 kHz 現場依存性大	
ロボット聴覚	音声認識率向上 音源方位推定	
カーナビ	音声認識率向上 音質向上	コスト
携帯電話	音質向上 低遅延化 ~ 数 ms	コスト 大きさ 消費電力
補聴器	音質向上 低遅延化 ~ 数 ms	コスト 大きさ 消費電力
マイク応用 (放送・計測等)	高周波帯域への拡張 ~ 48 kHz 音質向上 低遅延化 ~ 数 ms	

## 謝辞

本研究は株式会社神戸製鋼所技術開発本部および奈良先端科学技術大学院大学情報科学研究科音処理学講座にて行われました。

本研究を進めるにあたっては、本当に多くの方々にご協力頂きました。ここに深く感謝の意を表します。

本研究を進めるにあたり、社会人ドクターとして本学博士課程の受入れを快諾して頂き、また、ゼミでの議論をはじめ、貴重なご指導、ご助言を賜りました情報科学研究科 鹿野清宏教授に感謝致します。

本論文をまとめるにあたり、貴重なご指導、ご助言を賜りました情報科学研究科 杉本謙二教授に感謝致します。

本研究を進めるにあたり、2003年に始まった神戸製鋼所からの委託研究当時から常に相談に乗って頂き、熱心なご指導、ご助言を賜りました情報科学研究科 猿渡洋准教授に感謝致します。ここまで充実した研究ができたのも、猿渡先生との出会いがあつてこそと思います。本当にありがとうございました。

本研究を進めるにあたり、数多くのご討論、ご協力を賜りました情報科学研究科 森康充氏、高橋祐氏をはじめ情報科学研究科の皆様へ感謝致します。特に森康充氏は数年間共同研究を共にし、数多くの論文や学会発表を行ってきました。志半ばで逝去された時には本当に驚き、その事実を受け入れなければならない事は残念で仕方ないですが、その志は私の中にもしっかりと受け継いで今後の研究開発に生かしていきたいと思います。

本研究を進めるにあたり、本学博士課程への進学、さらには学費の会社負担まで快諾頂きました株式会社神戸製鋼所生産システム研究所 高橋哲也所長をはじめ研究所幹部の皆様へ感謝致します。

本研究を進めるにあたり、日々研究開発の進め方を議論し、共同で研究開発を行った株式会社神戸製鋼所生産システム研究所制御技術研究室 森田孝司室長、池田陽平氏、橋本裕志氏、山下俊郎氏、株式会社楓（フォン）張若愚氏に感謝致します。

本研究を進めるにあたり、音源分離モジュール、マイクロホン開発で多大なご協力を頂いた株式会社次世代研究所の皆様へ感謝致します。

最後に，本研究を進めるにあたり，日々の生活を支えてくれた家族に感謝致します．

## 参考文献

- [1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoustics, Speech, Signal Proc.*, vol.ASSP-27, no.2, pp.113–120, 1979.
- [2] G. W. Elko, "Microphone array system for hands-free telecommunication," *Speech Communication*, vol.20, pp.229–240, 1972.
- [3] O. Frost, "An algorithm for linearly constrained adaptive array processing," *Proceedings of the IEEE*, vol.60, pp.926–935, 1972.
- [4] R. Schmidt, "Multiple emitter location and signal parameter estimation," *Proceedings IEEE Transaction on Antennas and Propagation*, vol.34, pp.276–280, 1986.
- [5] J. F. Cardoso, "Eigenstructure of the 4th-order cumulant tensor with application to the blind source separation problem," *Proc. ICASSP'89*, vol.34, pp.2109–2112, 1989.
- [6] C. Jutten and J. Herault, "Blind separation of sources part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol.24, pp.1–10, 1991.
- [7] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol.36, pp.287–314, 1994.
- [8] A. J. Bell and T. J. Sejnowski, "An Information-Maximization Approach to Blind Separation and Blind Deconvolution," *Signal Processing*, vol.7, no.6, pp.1129–1159, 1995.
- [9] S. Amari, S. Douglas, A. Cichocki and H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," *Proceedings IEEE International Workshop on Wireless Communication*, pp.101–104, 1997.

- [10] P. Smaragdis, “Blind separation of convolved mixtures in the frequency domain,” *Neurocomputing*, vol.22, pp.21–34, 1998.
- [11] S. Choi, S. Amari, A. Cichocki and R. Liu, “Natural gradient learning with a nonholonomic constraint for blind deconvolution of multiple channels,” *Proceedings of International Workshop on ICA and BSS*, pp.371–376, 1999.
- [12] N. Murata and S. Ikeda, “An on-line algorithm for blind source separation on speech signals,” *Proc. NOLTA98*, vol.3, pp.923–926, 1998.
- [13] T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, “High-fidelity blind separation of acoustic signals using SIMO-model-based ICA with information-geometric learning,” *Proc. IWAENC2003*, pp.251–254, 2003.
- [14] T. Nishikawa, H. Saruwatari and K. Shikano, “Stable learning algorithm for blind separation of temporally correlated acoustic signals combining multi-stage ICA and linear prediction,” *IEICE Trans. Fundamentals*, vol.E86-A, no.8, pp.2028–2036, 2003.
- [15] L. Parra and C. Spence, “Convolutive blind separation of non-stationary sources,” *IEEE Trans. Speech & Audio Process.*, vol.8, pp.320–327, 2000.
- [16] H. Saruwatari, S. Kurita, K. Takeda, F. Itakura, T. Nishikawa, and K. Shikano, “Blind source separation combining independent component analysis and beamforming,” *EURASIP Journal on Applied Signal Processing*, vol.2003, pp.1135–1146, 2003.
- [17] T. Nishikawa, H. Saruwatari and K. Shikano, “Blind source separation of acoustic signals based on multistage ICA combining frequency-domain ICA and time-domain ICA,” *IEICE Trans. Fundamentals*, vol.2003, vol.E86-A, no.4, pp.846–858, 2003.
- [18] H. Sawada, R. Mukai, S. Araki, and S. Makino, “Polar coordinate based nonlinear function for frequency domain blind source separation,” *IEICE*

- Trans. Fundamentals*, vol.E86-A, no.3, pp.590–596, 2003.
- [19] S. Araki, S. Makino, Y. Hinamoto, R. Mukai, T. Nishikawa and H. Saruwatari, “Equivalence between Frequency Domain Blind Source Separation and Frequency Domain Adaptive Beamforming for Convulsive Mixtures,” *EURASIP Journal on Applied Signal Processing*, vol.2003, no.11, pp.1157–1166, 2003.
- [20] T. Takatani, T. Nishikawa, H. Saruwatari and K. Shikano, “High-fidelity blind separation of acoustic signals using SIMO-model-based independent component analysis,” *IEICE Trans. Fundamentals*, vol.E87-A, no.8, pp.2063–2072, 2004.
- [21] T. Takatani, S. Ukai, T. Nishikawa, H. Saruwatari and K. Shikano, “A self-generator method for initial filters of SIMO-ICA applied to blind separation of binaural sound mixtures,” *IEICE Trans. Fundamentals*, vol.E88-A, no.7, pp.1673–1682, 2005.
- [22] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda and F. Itakura, “Evaluation of blind signal separation method using directivity pattern under reverberant conditions,” *Proceedings of 2000 International Conference on Acoustics, Speech, and Signal Processing (ICASSP2000)*, vol.5, pp.3140–3143, 2000.
- [23] H. Sawada, R. Mukai, S. Araki and S. Makino, “A robust and precise method for solving the permutation problem of frequency-domain blind source separation,” *Proceeding of International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pp.505–510, 2003.
- [24] H. Saruwatari, T. Kawamura, T. Nishikawa and K. Shikano, “Fast-convergence algorithm for blind source separation based on array signal processing,” *IEICE Trans. Fundamentals*, vol.E86-A, no.4, pp.286–291, 2003.
- [25] A. Poularikas, “The Handbook of Formulas and Tables for Signal Processing,” *CRC Press, Boca Raton, Fla.*, 1999.

- [26] H. Buchner, R. Aichner, and W. Kellermann, “A generalization of blind source separation algorithms for convolutive mixtures based on second order statistics,” *IEEE Trans. Speech Audio Processing*, vol.13, no.1, pp.120–134, 2005.
- [27] Y. Mori, H. Saruwatari, T. Takatani, S. Ukai, K. Shikano, T. Hiekata, Y. Ikeda, H. Hashimoto, and T. Morita, “Blind separation of acoustic signals combining SIMO-model-based independent component analysis and binary masking,” *EURASIP Journal on Applied Signal Processing*, vol.2006, Article ID 34970, 17 pages, 2006.
- [28] Y. Mori, H. Saruwatari, T. Takatani, S. Ukai, K. Shikano, T. Hiekata, Y. Ikeda, H. Hashimoto, and T. Morita, “Blind separation of acoustic signals combining SIMO-model-based independent component analysis and binary masking,” *EURASIP Journal on Applied Signal Processing*, vol.2006, Article ID 34970, 17 pages, 2006.
- [29] Y. Takahashi, T. Takatani, H. Saruwatari and K. Shikano, “Blind spatial subtraction array with independent component analysis for hands-free speech recognition,” *Proc. of IWAENC*, 2006.
- [30] D. Kolossa and R. Orglmeister, “Nonlinear Postprocessing for Blind Speech Separation,” *Proceeding of International Workshop on Independent Component Analysis and Blind Signal Separation (ICA2004)*, pp.832-839, 2004.
- [31] R. Lyon, “A computational model of binaural localization and separation,” *Proc. ICASSP83*, pp.1148–1151, 1983.
- [32] N. Roman, D. Wang and G. Brown, “Speech segregation based on sound localization,” *Proc. IJCNN01*, pp.2861–2866, 2001.
- [33] M. Aoki, M. Okamoto, S. Aoki, H. Matsui, T. Sakurai, and Y. Kaneda, “Sound source segregation based on estimating incident angle of each fre-

- quency component of input signals acquired by multiple microphones,” *Acoustical Science and Technology*, vol.22, no.2, pp.149–157, 2001.
- [34] Y. Mori, H. Saruwatari, K. Shikano, T. Hiekata, and T. Morita, “Directivity-dependency-reduced blind source separation integrating ICA, beamforming and binary masking,” *Proc. IROS2007*, 2007.
- [35] L. Rabiner and B. Juang, “Fundamentals of speech recognition,” *Prentice Hall PTR, Upper Saddle River, NJ*, 1993.
- [36] “Mapping function for transforming P862 raw result scores to MOS-LQO,” *ITU-T P.862.1*, 2003.
- [37] T. Kobayashi, S. Itabashi, S. Hayashi and T. Takezawa, “ASJ continuous speech corpus for research,” *The Journal of The Acoustic Society of Japan*, vol.42, no.12, pp.888–893, 1992.
- [38] 稗方孝之, 池田陽平, 橋本裕志, 森田孝司, 森康充, 高谷智哉, 猿渡洋, 鹿野清宏, “SIMO-ICA と SIMO バイナリマスクに基づく実時間音源分離モジュールの開発,” *日本音響学会講演論文集*, 3-Q-5, pp.637–638, March 2006.
- [39] T. Hiekata, Y. Ikeda, T. Yamashita, T. Morita, R. Zhang, Y. Mori, H. Saruwatari, and K. Shikano, “Development and evaluation of pocket-size real-time blind source separation microphone,” *Proc. of IWAENC*, 2008.
- [40] 猿渡洋, “音声・音響信号を対象としたブライント音源分離,” *電子情報通信学会 DSP 研究会*, DSP2001-194, pp.59–66, March 2002.
- [41] 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ト音源分離,” *電子情報通信学会技術報告*, EA2003-45, pp.17–24 June 2003.
- [42] Y. Takahashi, K. Osako, H. Saruwatari and K. Shikano, “Blind source extraction for hands-free speech recognition based on wiener filtering and ICA-based noise estimation,” *Proc. of HSCMA*, pp.164–167, 2008.



- [43] 大迫慶一, 高橋祐, 森康充, 猿渡洋, 鹿野清宏, "独立成分分析に基づく近接点音源除去の検討," 日本音響学会講演論文集, 1-7-16, pp.647-650, Sept. 2007.
- [44] S. Ikeda and N. Murata, "A method of ICA in the frequency domain," *Proc. Intern. Workshop on ICA and BSS*, pp.365-371, 1999.
- [45] M. Brandstein and D. Ward, Eds., "Microphone Arrays: Signal Processing Techniques and Applications," *Springer-Verlag*, 2001.
- [46] "Japanese speech corpus for large vocabulary continuous speech recognition research," *The Journal of The Acoustic Society of Japan (E)*, pp.199-206, 1999.
- [47] "A new phonetic tied-mixture model for efficient decoding," pp.1269-1272, 2000.
- [48] "Julius - an open source real-time large vocabulary recognition engine," pp.1691-1694, 2001.

## 付録

### A. ICA による拡散性雑音推定の理論的考察と実験

厳密な証明は本論文の範囲を超えるので行わないが，読者の理解を助けるために，拡散性雑音環境でのICAの振舞いについて述べておく．詳細については既発表文献 [42] を参照されたい．

#### A.1 ICA により形成されるフィルタの具体例

ここでは，実際にICAによってどのような特性を持つフィルタが形成されるかを検討する．想定する環境のモデルは，空間的無相関雑音下において近接点音源である話者が1人いる場合である．すなわちこれは，携帯電話の音声通話やロボット・カーナビの音声認識システムなど，実環境を想定したモデルである．

まず，点音源である目的音声を抽出するビームフォーマ  $g_s(f)$  について検討する．所望のビームフォーマは，目的音声の到来方位 (Direction of arrival: DOA) を与えたもとで，非目的音のパワーを最小化するものであり，これはMVDR (Minimum variance distortionless response) ビームフォーマである [19, 45]．このビームフォーマ  $g_s(f)$  は下式のように表される．

$$g_s^T(f) = \frac{\mathbf{a}(f, \theta_s(f))^H \mathbf{R}^{-1}(f)}{\mathbf{a}(f, \theta_s(f))^H \mathbf{R}^{-1}(f) \mathbf{a}(f, \theta_s(f))} \quad (44)$$

$$\mathbf{a}(f, \theta_s(f)) = [\exp(i2\pi(f/M)f_s d_1 \sin \theta_s/c), \dots, \exp(i2\pi(f/M)f_s d_J \sin \theta_s/c)]^T \quad (45)$$

ここで， $\mathbf{a}(f, \theta_s(f))$  はステアリングベクトル， $\theta_s(f)$  はターゲット音声のDOAである． $M$  はDFTのフィルタ長， $f_s$  はサンプリング周波数， $d_j$  はマイクロホンの座標， $c$  は音速である．そして  $\mathbf{R}(f)$  は， $\mathbf{n}(f, \tau)$  の相関行列であり，

$$\mathbf{R}(f) = \langle \mathbf{n}(f, \tau) \mathbf{n}^H(f, \tau) \rangle_\tau \quad (46)$$

である．厳密には，低周波数帯域において相関を持つが [45]，ここでは便宜上，全ての周波数帯域において無相関になると仮定する．雑音のパワーを  $\delta^2(f)$  とする

と、相関行列は  $R(f) = \delta^2(f) \cdot I$  で与えられる。ここで、 $I$  は単位行列である。これより、相関行列の逆行列は  $R^{-1}(f) = I/\delta^2(f)$  となり、式 (44) に代入すると以下ようになる。

$$\mathbf{g}_s^\top(f) = \frac{\mathbf{a}(f, \theta_s(f))^\text{H}}{\mathbf{a}(f, \theta_s(f))^\text{H} \mathbf{a}(f, \theta_s(f))} \quad (47)$$

ここで、分母の項は  $\mathbf{a}(f, \theta_s(f))^\text{H} \mathbf{a}(f, \theta_s(f)) = J$  であるから、最終的に  $\mathbf{g}_s(f)$  は、

$$\begin{aligned} \mathbf{g}_s(f) = & \frac{1}{J} [\exp(-i2\pi(f/M)f_s d_1 \sin \theta_s(f)/c), \\ & \dots, \exp(-i2\pi(f/M)f_s d_J \sin \theta_s(f)/c)]^\text{T} \end{aligned} \quad (48)$$

となり、これはすなわち遅延和アレー (Delay-and-sum: DS) である。ただし、この  $\mathbf{g}_s(f)$  は通常の DS ではなく、各周波数帯域でターゲット音声の DOA である  $\theta_s(f)$  に指向特性が形成されており、残響に適應した DS である。DS の性能はマイクロホンの素子数に比例し、その SNR は  $10 \log_{10} J$  [dB] で与えられる。従って、ターゲット音声のみを抽出するには、非常に大規模なマイクロホンアレーを必要とする。

次に、雑音を推定するビームフォーマ  $\mathbf{g}_n(f)$  について検討する。雑音推定は、ターゲット音声である近接点音源を除去するものであり、 $\theta_s(f)$  方位に死角を向けた死角制御型ビームフォーマ (Null beamforming: NBF) が SNR 最大化ビームフォーマとなる。例えば 2 素子の場合、NBF は以下の式で定義される。

$$\begin{aligned} \mathbf{g}_n(f) = & [\exp(-i2\pi(f/M)f_s d_1 \sin \theta_s(f)/c), \\ & - \exp(-i2\pi(f/M)f_s d_J \sin \theta_s(f)/c)]^\text{T} \cdot \sigma(f) \end{aligned} \quad (49)$$

ここで、 $\sigma(f)$  は利得の補正係数である。DS と同様にこの NBF は、各周波数帯域で各々の方位  $\theta_s(f)$  に死角が形成されるため、残響に適應した NBF である。従って、近接点音源を除去する性能は極めて高いと推測できる。

以上をまとめると、空間的無相関雑音下に存在する近接点音源を分離する場合、ICA の学習によって得られる SNR 最大化ビームフォーマは以下ようになる。

- 近接点音源を抽出するビームフォーマ  $\mathbf{g}_s(f)$  は DS となる。性能は高くなく、ターゲット音声のみを抽出することには限界がある。

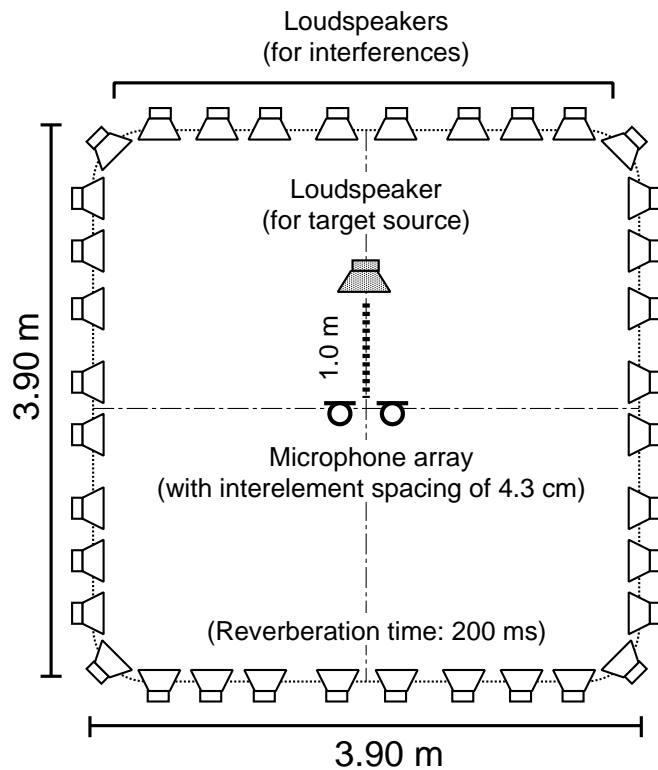


図 39 実験に用いた残響部屋のレイアウト

- 雑音を抽出する，つまり近接点音源を除去するビームフォーマ  $g_n(f)$  は NBF となる．性能は非常に高く，高精度な雑音推定を行うことができる．

## A.2 実験条件

本実験は，上記理論を確認するために行った．実験を行った部屋を図 39 に示す．マイクロホンアレーの正面に目的音声，その周囲に雑音源を配置した．目的音声は，ラウドスピーカからマイクロホンまでのインパルス応答とドライソースを畳み込んで作成している．ドライソースの音声は男性 6 名，女性 6 名の計 12 名である．雑音は，周囲の 36 個のラウドスピーカからそれぞれ異なる信号を出力し，それらを同時にマイクロホンで収録した．出力する信号は，白色雑音，駅で

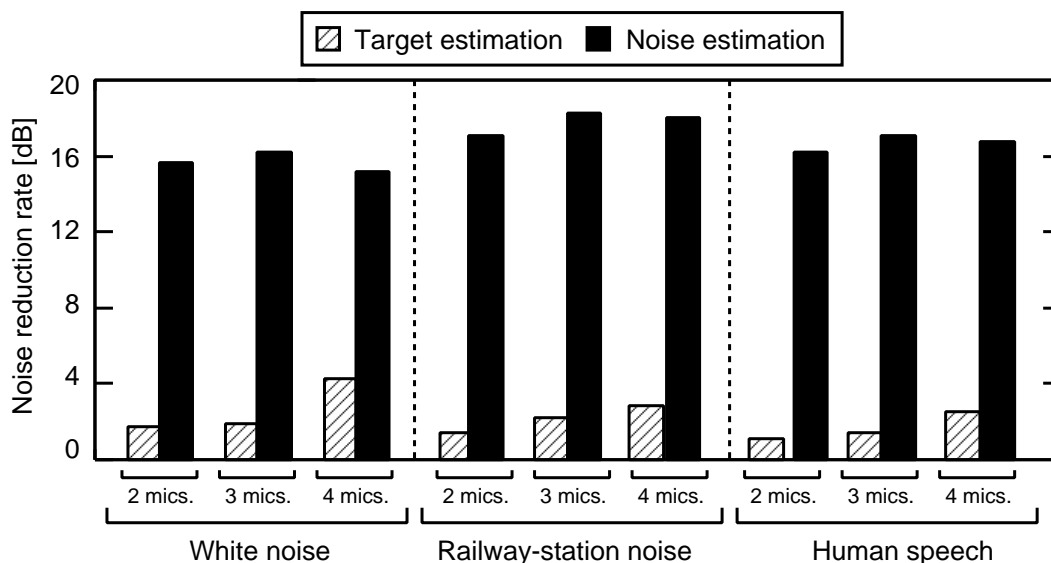


図 40 分離実験のシミュレーション結果

収録した雑音，話者が全て異なる音声の3種類である．これらの信号のサンプリング周波数は8 kHzである．部屋の残響時間は200 ms，マイクロホンの素子数は2, 3, 4であり間隔はそれぞれ4.3 cmとした．

### A.3 実験結果

各雑音環境下でのNRRを図40に示す．白色雑音，駅雑音，音声雑音の全てにおいて雑音推定のNRRが高く，一方でターゲット音声推定はさほど高くないことがわかる．

次に図41は，ICAの学習によって形成された分離フィルタの指向特性を示したものである．実線はターゲット音声推定の指向特性，破線は雑音推定の指向特性である．ターゲット音声推定の指向特性を見ると， $0^\circ$ 方位にビームが向けられており，このビームフォーマはDSとなっている事が分かる．また，雑音推定の指向特性を見ると，こちらもほぼ $0^\circ$ 方位に鋭い死角が形成されており，このビームフォーマはNBFである事が分かる．

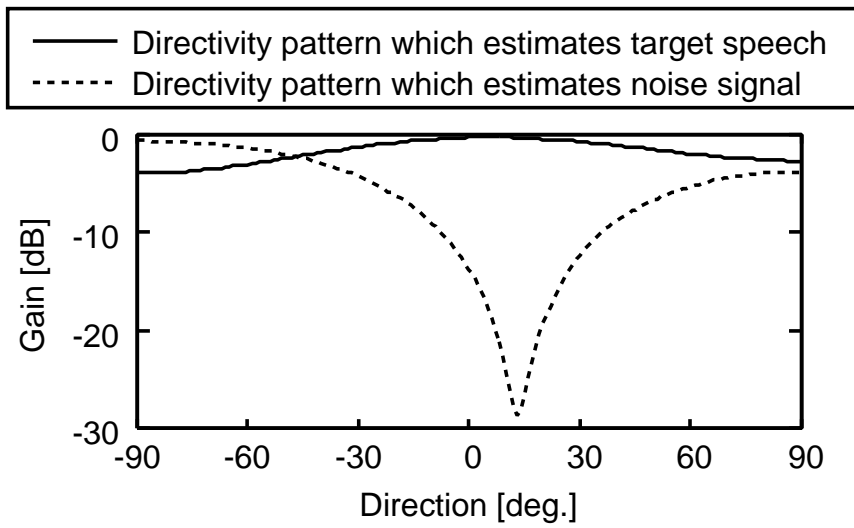


図 41 空間的無相関雑音環境下における 1.5 kHz かつ 2 素子アレーの場合の典型的な ICA フィルタの指向パターン

## B. SIMO-ICA と SIMO-BM を結合した 2 段階 BSS 法を用いた音声認識実験

読者の理解を助けるために，BSS マイクロホンに実装されている SIMO-ICA と SIMO-BM を結合した 2 段階 BSS 法を用いた大語彙音声認識実験結果を示す．なお，詳細については既発表文献 [27] を参照されたい．

### B.1 実験条件

表 8 に音声認識実験条件を示す．音源 1 ( $S_1(f)$ ) としてテストセットの 200 文を，音源 2 ( $S_2(f)$ ) には干渉音としてテストセットとは異なる 1 文章を用い，混合信号の SNR は 0 dB である．音源分離は混合音から音源 1 を取り出す機能として働き，それを音声認識に用いる．

表 8 音声認識実験のパラメータ

Database	JNAS [46], 306 speakers (150 sentences / speaker)
Task	20-k newspaper dictation
Acoustic model	phonetic tied mixture [47] (clean model)
Feature vectors	12-order MFCCs, 12-order $\Delta$ MFCCs, 1-order $\Delta$ energy
Training data	260 speakers' utterances (150 sentences / speaker)
Testing data	46 speakers' utterances (200 sentences)
Decoder	Julius [48] ver.3.4.2
Sampling frequency	16 kHz
Frame length	25 ms
Frame shift	10 ms

## B.2 実験結果

図 42 に各音源方位組み合わせ・各手法ごとの単語認識精度 (WA) の結果を示す。これによれば、提案法 (BSS マイクロホンに実装されている音源分離法) は、バイナリマスク, ICA, およびバイナリマスクと ICA の単純接続と比較して性能が良いことが分かる。この結果から、音声認識においては、雑音抑圧性能に加えて、高音質性能 (歪みが少ないこと) が音声認識に有効であることを示している。BSS マイクロホンはこの両方の性能を備えたリアルタイム装置であり、雑音環境下でのリアルタイム音声認識への応用が期待できる。

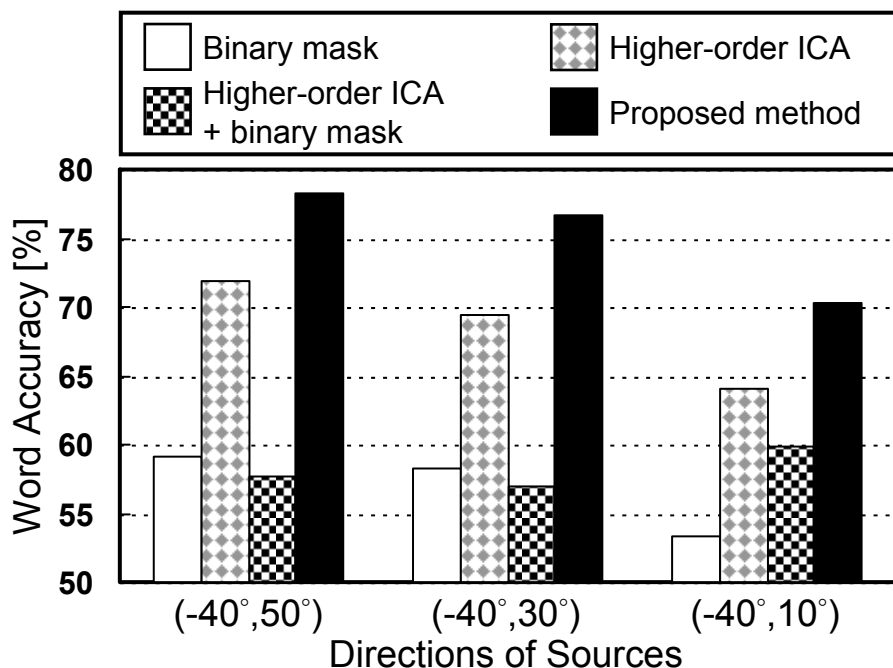


図 42 話者位置と手法の違いによる単語認識精度の評価結果。音声認識は、2 万語の新聞読み上げ認識である。音声デコーダは Julius [48] のバージョン 3.4.2 を用い、音響モデルは JNAS [46] データベースから選択した 260 話者 (1 話者あたり 150 文を読み上げ) を学習に用いた PTM モデル [47] を用いている。



## 発表リスト

### 学術論文

1. Takashi Hiekata, Yohei Ikeda, Toshiro Yamashita, Takashi Morita, Ruoyu Zhang, Yoshimitsu Mori, Hiroshi Saruwatari, and Kiyohiro Shikano, “Development and evaluation of pocket-size real-time blind source separation microphone,” *Acoustical Science and Technology*, Vol.30, No.4, pp.297–304, July 2009. (第3章に関連)
2. 稗方孝之, 池田陽平, 橋本裕志, 森田孝司, 張若愚, 高橋祐, 猿渡洋, 鹿野清宏, “並列ICAによる雑音推定に基づくリアルタイムブラインド音源抽出マイクロホンの開発”, 電子情報通信学会誌, Vol.J92-D, No.10, pp.1772–1783, Oct. 2009. (第4章に関連)
3. Yoshimitsu Mori, Hiroshi Saruwatari, Tomoya Takatani, Satoshi Ukai, Kiyohiro Shikano, Takashi Hiekata, Youhei Ikeda, Hiroshi Hashimoto and Takashi Morita, “Blind separation of acoustic signals combining SIMO-model-based independent component analysis and binary masking,” *EURASIP Journal on Applied Signal Processing*, vol.2006, Article ID 34970, 17 pages, 2006.

### 招待講演

1. 稗方孝之, 池田陽平, 森田孝司, “リアルタイム音源分離技術の開発と実用化へ向けた取組み”, 第51回システム制御情報学会研究発表講演会講演論文集, pp.7–12, May 2007.

### 解説

1. 猿渡洋, 森康充, 高谷智哉, 鹿野清宏, 稗方孝之, 森田孝司, “単一入力・多点出力系に基づく独立成分分析を用いた実時間音源分離,” 計測と制御, vol.46, no.6, pp.434-440, 2007.

## 国際会議

1. Takashi Hiekata, Youhei Ikeda, Toshiro Yamashita, Takashi Morita, Ruoyu Zhang, Yoshimitsu Mori, Hiroshi Saruwatari, Kiyohiro Shikano, “Development and evaluation of pocket-size real-time blind source separation microphone,” *Proceedings of 2008 International Workshop on Acoustic Echo and Noise Control (IWAENC2008)*, Sept. 2008.
2. Takashi Hiekata, Takashi Morita, Youhei Ikeda, Hiroshi Hashimoto, Ruoyu Zhang, Yuu Takahashi, Hiroshi Saruwatari, Kiyohiro Shikano, “Multiple ICA-based real-time blind source extraction applied to handy size microphone,” *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2009)*, pp.121–124, April 2009.
3. Hiroshi Saruwatari, Yoshimitsu Mori, Tomoya Takatani, Satoshi Ukai, Kiyohiro Shikano, Takashi Hiekata, Takashi Morita, ”Two-Stage Blind Source Separation Using SIMO-ICA and Binary Masking,” *Proceedings of Joint Workshop on Hands-Free Speech Communication and Microphone Arrays (HSCMA2005)*, pp.d-11–12, March 2005.
4. Yoshimitsu Mori, Tomoya Takatani, Satoshi Ukai, Hiroshi Saruwatari, Kiyohiro Shikano, Takashi Hiekata, Takashi Morita, “Blind Separation of Convolutional Speech Mixtures Using SIMO-Model-Based ICA and Binary Mask Processing,” *Proceedings of International Workshop on Nonlinear Signal and Image Processing (NSIP2005)*, pp.187–192, May 2005.
5. Hiroshi Saruwatari, Yoshimitsu Mori, Tomoya Takatani, Satoshi Ukai, Kiyohiro Shikano, Takashi Hiekata, Takashi Morita, ”Two-Stage Blind Source Separation Based on ICA and Binary Masking for Real-Time Robot Audition System,” *Proceedings of International Conference on Intelligent Robots and Systems (IROS2005)*, pp.215–220, August 2005.

6. Yoshimitsu Mori, Hiroshi Saruwatari, Tomoya Takatani, Satoshi Ukai, Kiyohiro Shikano, Takashi Hiekata, Takashi Morita, “Real-Time Implementation of Two-Stage Blind Source Separation Combining SIMO-ICA and Binary Masking,” *Proceedings of 2005 International Workshop on Acoustic Echo and Noise Control (IWAENC2005)*, pp.229–232, Sept. 2005.
7. Yoshimitsu Mori, Tomoya Takatani, Hiroshi Saruwatari, Kiyohiro Shikano, Takashi Hiekata, Takashi Morita, “ICA and Binary-Mask-Based Blind Source Separation with Small Directional Microphones,” *International Conference on Independent Component Analysis and Blind Source Separation (ICA2006)*, LNCS 3889, pp. 649–657, March 2006.
8. Yoshimitsu Mori, Tomoya Takatani, Hiroshi Saruwatari, Kiyohiro Shikano, Takashi Hiekata, Takashi Morita, “Blind Source Separation Combining SIMO-ICA and SIMO-Model-Based Binary Masking”, *International Conference on Acoustics, Speech, and Signal Processing (ICASSP2006)*, pp.V-81–84, May 2006.
9. Yoshimitsu Mori, Tomoya Takatani, Hiroshi Saruwatari, Kiyohiro Shikano, Takashi Hiekata, Takashi Morita, “Two-Stage Blind Separation of Moving Sound Sources with Pocket-Size Real-Time DSP Module”, *the 14th European Signal Processing Conference (EUSIPCO2006)*, Sept. 2006 (accepted).
10. Hiroshi Saruwatari, Yoshimitsu Mori, Tomoya Takatani, Kiyohiro Shikano, Takashi Hiekata, Youhei Ikeda, Hiroshi Hashimoto, Takashi Morita, “Real-Time Blind Separation of Acoustic Signals Using SIMO-Model-Based Independent Component Analysis,” *Joint 3rd International Conference on Soft Computing and Intelligent Systems and 7th International Symposium on advanced Intelligent Systems (SCIS & ISIS 2006)*, Sept. 2006 (Invited Talk).
11. Yoshimitsu Mori, Tomoya Takatani, Hiroshi Saruwatari, Kiyohiro Shikano,

Takashi Hiekata, Takashi Morita, “Noise-robust Hands-free Speech Recognition Using SIMO-model Based Blind Source Separation,” *International Symposium on Signal Processing and its Application (ISSPA2007)*, February 2007 (CD-ROM) (Invited Talk).

12. Yoshimitsu Mori, Tomoya Takatani, Hiroshi Saruwatari, Kirohiro Shikano, Takashi Hiekata, Takashi Morita, “Evaluation of Blind Source Separation Combining SIMO-ICA and SIMO-Model-Based Binary Masking in Noisy Environment,” *2007 RISP International Workshop on Nonlinear Circuits and Signal Processing (NCSP 07)*, pp.225–228, March 2007.
13. Yoshimitsu Mori, Tomoya Takatani, Hiroshi Saruwatari, Kiyohiro Shikano, Takashi Hiekata, Takashi Morita, “High-Presence Hearing-Aid System Using DSP-Based Real-Time Blind Source Separation Module,” *International Conference on Acoustics, Speech, and Signal Processing (ICASSP2007)*, Vol.4, pp.609–612, April 2007.
14. Yoshimitsu Mori, Hiroshi Saruwatari, Kiyohiro Shikano, Takashi Hiekata, Takashi Morita, “Directivity-Dependency-Reduced Blind Source Separation Integrating ICA, Beamforming and Binary Masking,” *2007 IEEE International Conference on Intelligent Robots and Systems (IROS2007)*, Oct. 2007 (Invited Talk).

## 研究会

1. Takashi Hiekata, Youhei Ikeda, Hiroshi Hashimoto, Takashi Morita, Ruoyu Zhang. Zhang, Yuu Takahashi, Hiroshi Saruwatari, Kiyohiro Shikano, “Development of blind signal extraction device based on noise estimation by parallel ICAs,” *IEICE Technical Report*, EA2008-135, pp.115–120, January 2009.

2. Yoshimitsu Mori, Tomoya Takatani, Satoshi Ukai, Hiroshi Saruwatari, Kiyohiro Shikano, Takashi Hiekata, Takashi Morita, "Blind source separation combining SIMO-model-based ICA and binary mask processing," *IEICE Technical Report*, EA2004-115, pp.71–76, December 2004.
3. 森康充, 高谷智哉, 猿渡洋, 鹿野清宏, 稗方孝之, 森田孝司, "SIMO-ICA とバイナリマスク処理を組み合わせた 2 段型ブラインド音源分離", *AI チャレンジ研究会*, October 2005.
4. 森康充, 高谷智哉, 猿渡洋, 鹿野清宏, 稗方孝之, 森田孝司, "SIMO-ICA とバイナリマスク処理を組み合わせた 2 段リアルタイム音源分離モジュールの開発", 日本音響学会関西支部第 8 回若手研究者交流研究発表会, p.12, December 2005.
5. 森康充, 高谷智哉, 猿渡洋, 鹿野清宏, 稗方孝之, 森田孝司, "SIMO-ICA と SIMO モデルに基づくバイナリマスク処理を結合した歪の少ないブラインド音源分離", 電子情報通信学会技術報告, EA2005-102, pp. 37–42, January 2006

## 大会発表

1. 稗方孝之, 池田陽平, 橋本裕志, 森田孝司, 森康充, 高谷智哉, 猿渡洋, 鹿野清宏, "SIMO-ICA と SIMO バイナリマスクに基づく実時間ブラインド音源分離モジュールの開発", 日本音響学会講演論文集, 3-Q-5, pp.637-638, March 2006.
2. 稗方孝之, 池田陽平, 山下俊郎, 森田孝司, 森康充, 猿渡洋, 鹿野清宏, "ポケットサイズ音源分離マイクロホン SSM-001 の開発", 日本音響学会講演論文集, 1-P-13, March 2008.
3. 稗方孝之, 池田陽平, 橋本裕志, 森田孝司, 張若愚, 高橋祐, 猿渡洋, 鹿野清宏, "複数 ICA を用いた雑音推定に基づくブラインド音源抽出マイクロホンの開発", 日本音響学会講演論文集, 3-P-29, March 2009.

4. 森 康充, 高谷 智哉, 鵜飼 訓史, 猿渡 洋, 鹿野 清宏, 稗方 孝之, 森田 孝司, "SIMO モデルに基づく ICA とバイナリマスク処理を用いた 2 段ブラインド音源分離手法", 日本音響学会講演論文集, 1-6-12, pp.431-432, March 2005.
5. 森 康充, 高谷 智哉, 鵜飼 訓史, 猿渡 洋, 鹿野 清宏, 稗方 孝之, 森田 孝司, "SIMO-ICA とバイナリマスク処理を組み合わせたリアルタイムブラインド音源分離手法の評価", 日本音響学会講演論文集, 3-Q-27, pp. 701-702, Sept. 2005.
6. 森 康充, 高谷 智哉, 猿渡 洋, 鹿野 清宏, 稗方 孝之, 森田 孝司, "SIMO-ICA と SIMO モデルに基づく拡張バイナリマスク処理を用いた 2 段ブラインド音源分離手法", 日本音響学会講演論文集, 1-5-16, pp.537-538, March 2006.
7. 森 康充, 高谷 智哉, 三橋 禎, 猿渡 洋, 鹿野 清宏, 稗方 孝之, 森田 孝司, "リアルタイムブラインド音源分離モジュールを用いた高臨場補聴システム," 日本音響学会講演論文集, 3-Q-10, pp.519-520, Sept. 2006.
8. 森 康充, 高谷 智哉, 猿渡 洋, 鹿野 清宏, 稗方 孝之, 森田 孝司, "ブラインド音源分離モジュールと聴覚補助システム応用," 平成 18 年電気関係学会関西支部連合大会, G16-14, p.G377, November 2006.
9. 森 康充, 高谷 智哉, 猿渡 洋, 鹿野 清宏, 稗方 孝之, 森田 孝司, "リアルタイムブラインド音源分離モジュールを用いたバイノーラル補聴システム," 第 9 回関西支部若手研究者交流研究発表会, p.12, Dec. 2006.
10. 森 康充, 高谷 智哉, 猿渡 洋, 鹿野 清宏, 稗方 孝之, 森田 孝司, "SIMO-ICA とバイナリマスク処理を用いた 2 段ブラインド音源分離における音源方位依存性の改善," 日本音響学会講演論文集, 3-P-9, pp.613-614, March 2007.

## 技術報告

1. 稗方 孝之, 森田 孝司, "リアルタイム音源分離技術" *R&D* 神戸製鋼技報, Vol.56, No.1, pp.76, April 2006.

## 展示会

1. 2006/03 第5回ケータイ国際フォーラム (奈良先端大と神戸製鋼共同展示)
2. 2006/09 イノベーション・ジャパン (奈良先端大と神戸製鋼共同展示)
3. 2007/11 システムコントロールフェア 2007 (ORiN 協議会ブース)
4. 2007/11 国際ロボット展 (ORiN 協議会ブース)

## 受賞

1. International Conference on Intelligent Robots and Systems(IROS2005) , Application Paper Award, "Two-Stage Blind Source Separation Based on ICA and Binary Masking for Real-Time Robot Audition System," October 2006.
2. 第24回電気通信普及財団テレコムシステム技術賞, "Blind Separation of Acoustic Signals Combining SIMO-Model-Based Independent Component Analysis and Binary Masking," March 2009.
3. 第17回日本音響学会技術開発賞, "ポケットサイズリアルタイム音源分離・抽出マイクロホンの開発," May 2009.

## 特許リスト

1. 特開 2006-154314  
名称「音源分離装置，音源分離プログラム及び音源分離方法」  
発明者：稗方 孝之 森田 孝司 猿渡 洋
2. 特開 2007-219479  
名称「音源分離装置，音源分離プログラム及び音源分離方法」  
発明者：稗方 孝之 森田 孝司 猿渡 洋 森 康充
3. 特開 2008-35259  
名称「音源分離装置，音源分離方法及び音源分離プログラム」  
発明者：稗方 孝之 池田 陽平
4. 特開 2008-219458  
名称「音源分離装置，音源分離プログラム及び音源分離方法」  
発明者：池田 陽平 稗方 孝之 森田 孝司 猿渡 洋 森 康充
5. 特許第 4336378 号  
名称「目的音抽出装置，目的音抽出プログラム，目的音抽出方法」  
発明者：稗方 孝之 森田 孝司 池田 陽平 下田 敏章
6. 特開 2008-295010  
名称「目的音抽出装置，目的音抽出プログラム，目的音抽出方法」  
発明者：稗方 孝之 森田 孝司 池田 陽平 下田 敏章
7. 特開 2008-295011  
名称「目的音抽出装置，目的音抽出プログラム，目的音抽出方法」  
発明者：稗方 孝之 森田 孝司 池田 陽平 下田 敏章
8. 特願 2008-330478  
名称「目的音抽出装置及び目的音抽出プログラム」  
発明者：稗方 孝之



9. 意匠第 1298101 号  
名称「音響情報入出力器」  
創作者：幾谷 万智子 稗方 孝之 山下 俊郎
10. US 出願 12/223,069  
名称”Sound Source Separation Apparatus and Sound Source Separation Method”  
発明者：Takashi Hiekata, Takashi Morita, Hiroshi Saruwatari, Yoshimitsu Mori
11. US 公開 US-2009-0012779-A1  
名称”Sound Source Separation Apparatus and Sound Source Separation Method”  
発明者：Takashi Hiekata, Youhei Ikeda, Takashi Morita, Hiroshi Saruwatari, Yoshimitsu Mori
12. US 公開 US-2008-0267423-A1  
名称”Object Sound Extraction Apparatus and Sound Extraction Method”  
発明者：Takashi Hiekata, Takashi Morita, Youhei Ikeda, Toshiaki Shimoda

## プレスリリースリスト

### 1. テレビ放映

- 2005/10/24 NHK「ニュース7」「ニュース10」
- 2005/10/24 テレビ東京「ワールドビジネスサテライト(トレンドたまご)」
- 2005/10/24 テレビ大阪「ニュース525」
- 2005/11/07 フジテレビ「めざましテレビ」
- 2005/11/14 日本テレビ「ズームイン!!SUPER」

### 2. ラジオ

- 2005/11/05 NHK ラジオ第一「ラジオ朝一番」
- 2005/11/22 NHK Radio Japan(国際放送)

### 3. 新聞, 雑誌

- 2005/10/25 全国主要5紙・神戸
- 2005/10/25 日経・日経産業・日刊工業・産業・鉄鋼・化学工業
- 2005/11/04 フジサンケイビジネスアイ(音の出る紙面)
- 2005/11/07 The Nikkei Weekly
- 2005/11 The Times
- 2005/11/15 週刊アスキー
- 2005/11/28 日経ビジネス
- 2005/12 日経ものづくり
- 2006/04 電子情報通信学会誌ニュース解説 (Vol.89, No.4, p.354)
- 2007/09/11 日経産業「どんとこい未来技術」