

## 帯域分割型 CSP 法に基づく話者位置推定法の検討

傳田 遊亀<sup>†</sup> 西浦 敬信<sup>††</sup> 河原 英紀<sup>†</sup> 入野 俊夫<sup>†</sup>

<sup>†</sup> 和歌山大学大学院 システム工学研究科

〒 640-8510 和歌山市栄谷 930

<sup>††</sup> 立命館大学 情報理工学部

〒 525-8577 草津市野路東 1-1-1

E-mail: †{s045064,kawahara,irino}@sys.wakayama-u.ac.jp, ††nishiura@is.ritsumeikai.ac.jp

あらまし 音声制御システムやテレビ会議システムなどにおいて、発話者から離れた位置にあるマイクロホンで発話者の音声を高品質に受信することは極めて重要である。発話者から離れた位置のマイクロホンで発話者の音声を高品質に受信する方法として、マイクロホンアレイによる音声強調法が提案されている。しかし、マイクロホンアレイを用いて音声を強調するためには発話者の位置を推定することが必要となる。このため、例えば CSP (Cross-power Spectrum Phase) 法に基づく発話者位置推定法が提案されている。しかし、CSP 法には高雑音環境下で話者位置推定性能が低下してしまうという問題がある。そこで本稿では、音声の平均スペクトルで重み付けされた帯域分割型 CSP 法を提案し、音声に特化した位置推定法を提案する。また、提案手法によって発話者位置を推定した後、マイクロホンアレイによって目的音声を強調した場合の音声認識性能も合わせて評価する。実環境における評価実験の結果、提案手法は従来法より高い話者位置推定性能を得られることを確認した。

キーワード 発話者位置推定, CSP 法, 帯域分割分析, マイクロホンアレイ, 音声認識

## A Study of Talker Localization Based on Subband CSP Analysis

Yuki DENDA<sup>†</sup>, Takanobu NISHIURA<sup>††</sup>, Hideki KAWAHARA<sup>†</sup>, and Toshio IRINO<sup>†</sup>

<sup>†</sup> Graduate School of Systems Engineering, Wakayama University

930 Sakaedani, Wakayama, 640-8510 Japan

<sup>††</sup> College of Information Science and Engineering, Ritsumeikan University

1-1-1 Nojihigashi, Kusatsu, 525-8577 Japan

E-mail: †{s045064,kawahara,irino}@sys.wakayama-u.ac.jp, ††nishiura@is.ritsumeikai.ac.jp

**Abstract** It is very important to capture distant-talking speech with high quality for voice-controlled systems or teleconferencing systems. A microphone array steering is an ideal candidate as an effective method for capturing distant talking speech with high quality. However, it requires to localize a target talker before capturing distant-talking speech. For this purpose, a talker localization method based on CSP (Cross-power Spectrum Phase) analysis has been proposed, for example. However, talker localization performance of the CSP analysis is degraded in higher noisy environments. To deal with this problem, in this paper, we propose a subband CSP analysis with weighting of average speech spectrum and we propose a specialized localization method for speech. In addition, we evaluate the ASR (Automatic Speech Recognition) performance when the microphone array steering is steered to the estimated talker direction by the proposed method. As a result of evaluation experiments in a real room, we confirmed that the proposed method provides better talker localization performance than the conventional method.

**Key words** Talker localization, CSP analysis, Subband analysis, Microphone array, Speech recognition.

## 1. はじめに

テレビ会議システムや音声制御システムにおいて、話者から離れた位置にあるマイクロホンで話者の音声を受音する場合、残響や背景雑音の影響により受音した話者の音声の歪みを受け、音質が低下するという問題がある。そこで、話者から離れた位置にあるマイクロホンでも話者の音声を高音質に受音する方法として、マイクロホンアレーが注目されており、盛んに研究が行われている。これらの研究では、マイクロホンアレーを用いて話者の方向に指向性を形成することにより高音質な音声を受音を実現している。マイクロホンアレーを用いて形成される指向性は大きく分けると、話者の方向に対して超指向性を形成する遅延和アレー [1] と雑音や残響の方向に対して死角を形成する適応形アレー [2], [3] に分類できる。両手法とも指向性雑音を効果的に抑圧できることから、近年では音声認識などに盛んに利用が検討されている [4]。

しかしマイクロホンアレーでは、発話者の方向に指向特性を形成して発話者の音声を高品質に受音するため、発話者の位置を推定することが必要となる。このため、例えば最小分散法 (MV : Minimum Variance) に基づく話者位置推定法 [5]、MUSIC (MULiple Signal Classification) 法に基づく話者位置推定法 [6]、CSP (Cross-power Spectrum Phase) 法に基づく話者位置推定法 [7], [8] などが提案されている。これらの話者位置推定法のうち、計算量の観点で考えた場合、話者位置を 2 ch のマイクロホン (マイクロホンペア) のみで推定できる CSP 法は、非常に有効な手法である。また、CSP 法の性能改善を目的として、CSP 法を単一のマイクロホンペアから複数のマイクロホンペアに拡張した CSP 係数加算法 [9] も提案されている。しかし、CSP 係数加算法は使用するマイクロホンペアの数を増やし、それぞれのマイクロホンペアで CSP 係数を計算する必要がある。このため、本稿では単一のマイクロホンペアのみを使用して、正確に話者位置を推定することを検討する。

CSP 法は、相互相関の一種である CSP 係数に基づいて受音信号の到来方向を推定する手法である。CSP 法には、受音信号のスペクトル特性に依存しない正確な到来方向推定が行えるという特長があるため [10]、発話者のみが存在する環境では正確に話者位置を推定できる。しかし、高雑音環境下においては CSP 法の到来方向推定性能が劣化してしまうという問題があった。また、CSP 法は受音信号の全周波数帯域を使用して CSP 係数を計算し、到来方向を推定するため、目的信号の周波数特性を考慮した分析周波数帯域分割や分析重み係数を導入することで到来方向推定性能が改善される可能性がある。そこで本稿では、目的音を音声に限定した帯域分割型 CSP 法を提案し、話者位置を正確に推定することを検討する。

本稿では、まず節 2. で CSP 法に基づく話者位置推定法とその問題点について説明する。その後、節 3. において提案手法について説明し、節 4. で実環境で行った評価実験について説明する。

## 2. CSP (Cross-power Spectrum Phase) 法に基づく話者位置推定法とその問題点

本稿では、図 1 に示すように、マイクロホンペアに対して方向  $\theta$  から目的音声の到来状況を仮定する。この時、マイクロホン  $M_l$  と  $M_r$  で受音した信号  $s_l(t)$  と  $s_r(t)$  を使用して、CSP (Cross-power Spectrum Phase) 係数、到来時間差 (TDOA : Time Delay Of Arrival) および到来方向 (DAO : Direction Of Arrival) を式 (1)、式 (2) および式 (3) によって求めることができる。

式 (1) では、マイクロホンペアで受音した信号をフーリエ変換して、振幅で正規化を行う。そして位相差を求めた後、逆フーリエ変換を行い CSP 係数を求めている。そして式 (2) により、CSP 係数が最大になる時間差 (相関が最も高い時間差) を求めることにより、到来時間差の推定を行う。その後、式 (3) によって時間から角度への変換を行い到来方向を推定する。

$$\text{CSP}(k) = \text{IDFT} \left[ \frac{\text{DFT}[s_l(t)] \cdot \text{DFT}[s_r(t)]^*}{|\text{DFT}[s_l(t)]| \cdot |\text{DFT}[s_r(t)]|} \right], \quad (1)$$

$$\tau = \underset{k}{\text{argmax}} (\text{CSP}(k)), \quad (2)$$

$$\theta = \cos^{-1} \left( \frac{c \cdot \tau}{d \cdot F_s} \right), \quad (3)$$

ここで  $t$  と  $k$  は離散時間、DFT[ $\cdot$ ] は離散フーリエ変換 (DFT : Discrete Fourier Transform)、IDFT[ $\cdot$ ] は逆 DFT、\* は複素共役、 $\tau$  は到来時間差、 $c$  は音速、 $d$  はマイクロホンペア間の距離、 $F_s$  はサンプリング周波数、 $\theta$  は到来方向を表す。

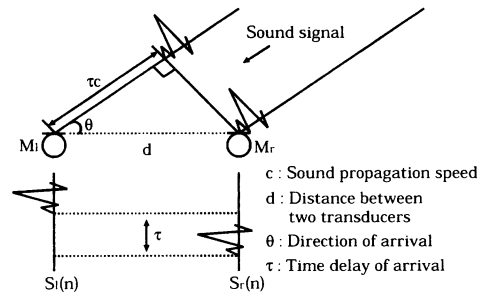


図 1 マイクロホンペアによる受音の原理

CSP 法は、式 (1) に示すように、受音信号の振幅で正規化された相互相関を使用することによって、受音信号のスペクトル特性に依存しない正確な到来方向推定を行うことができる [10]。このため、例えば図 2(a) に示すような目的音のみが存在する単一音源環境下では、図 2(b) に示すように、目的音の到来方向に対応する時間差の CSP 係数が高い値を持つため、CSP 法は目的音の到来方向を正確に推定することができる。ただし、図 2(b) では、CSP 係数を式 (4) により時間軸  $\text{CSP}(k)$  から方向軸  $\text{CSP}(\theta)$  に変換している。

$$\theta = \cos^{-1} \left( \frac{c \cdot k}{d \cdot F_s} \right), \quad (4)$$

しかし、CSP 法には到来方向推定性能が高雑音環境下で著しく劣化してしまうという問題点がある。図 3(a) に示すように、

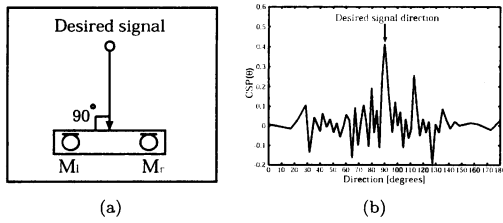


図2 目的音のみが存在する環境

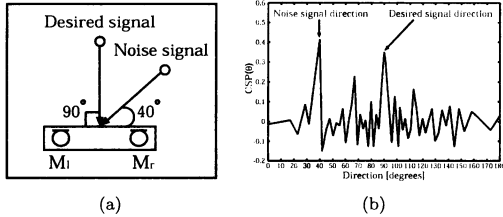


図3 目的音と雑音が存在する環境

マイクロホンペアが目的音のみでなく雑音も同時に受信するような環境においては、図3(b)に示すように、目的音の到来方向に対応する時間差のCSP係数よりも、雑音の到来方向に対応する時間差のCSP係数の方が高い値を持つことがある。この場合、CSP法は雑音の到来方向を目的音の到来方向として推定してしまうため、目的音の到来方向推定に失敗することになる。このような到来方向推定の失敗は、音声よりも雑音のパワーが大きい場合に発生する傾向がある。このため、高雑音環境下におけるCSP法の到来方向推定性能を改善するためには、目的音の到来方向のCSP係数を強調し、雑音の到来方向のCSP係数を抑圧する必要がある。そこで本稿では、CSP法の問題点について検討を行った。CSP法によって式(1)でCSP係数を計算する場合、受信信号の全周波数帯域を使用する。このため、周波数帯域分割を行い、目的信号のパワーが強い周波数帯域を使用してCSP係数を計算した場合、その周波数帯域では全周波数帯域での信号対雑音比(SNR: Signal to Noise Ratio)よりも相対的にSNRが向上することになるため、目的信号の到来方向のCSP係数が強調される可能性がある。また、目的信号の周波数特性に基づいた分析重み係数を導入し、各帯域のCSP係数に乗じることで、目的信号のパワーが強い周波数帯域で得られたCSP係数が強調され、さらに到来方向推定性能が改善される可能性がある。

一方、マイクロホンアレイを用いた音声認識では目的信号を人間の発話(音声)に限定することができる。そこで本稿では、雑音と音声の平均的なスペクトル特性は異

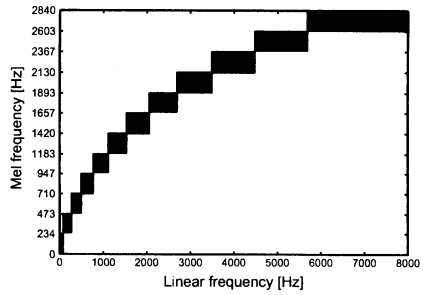


図5 周波数帯域分割の一例

なるという仮定に基づき、音声の平均スペクトルで重み付けされた帯域分割型CSP法を提案し、CSP法を話者位置位置推定に特化させることを検討する。

### 3. 提案手法

提案手法の概要を図4に示す。提案手法では、音声の特性を考慮した周波数帯域分割を行い、各周波数帯域ごとにCSP係数を計算する。次に、音声の平均スペクトルから学習した分析重み係数をCSP係数に乘じ、各周波数帯域のCSPを加算することで最終的なCSP係数を計算し、話者位置を推定する。

本稿では、節3.1で周波数帯域分割について、節3.2で分析重み係数について、節3.3で周波数帯域選択について、節3.4で話者位置推定について述べる。

#### 3.1 周波数帯域分割

人間の可聴範囲が20~20,000 Hzであるのに対して、音声はおおよそ100~6,000 Hzである。音声の短時間的性質として、第1フォルマント、第2フォルマントや第3フォルマントを含む低い周波数帯域に音声のパワーが集中する。また、音声の統計的性質として、音声の平均スペクトルは800 Hzまではほぼ平坦な傾斜を持ち、800 Hz以上は-10 dB/oct.程度の傾斜を持つ[11]。このことから、音声のパワーは低い周波数帯域ほど高いことがわかる。従って本稿では、低い周波数帯域を細かく分析し、高い周波数帯域は荒く分析を行うために、メル周波数上で等間隔に周波数帯域分割を行い、各帯域ごとにCSP係数を計算する。線形周波数  $f$  からメル周波数  $f_{Mel}$  への変換式を式(5)に示す。

$$f_{Mel} = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right), \quad (5)$$

ここで  $f_{Mel}$  はメル周波数、 $f$  は線形周波数を表す。

図5に周波数帯域分割の一例を示す。図5に示すように、メル周波数軸上で等間隔に帯域分割を行うことで、線形周波数軸

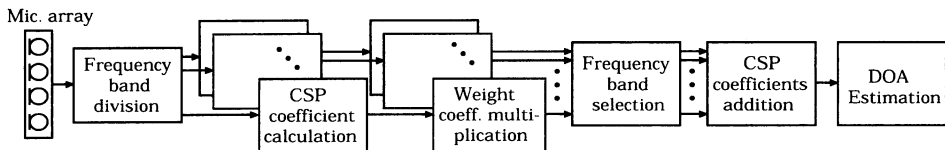


図4 提案手法の概要

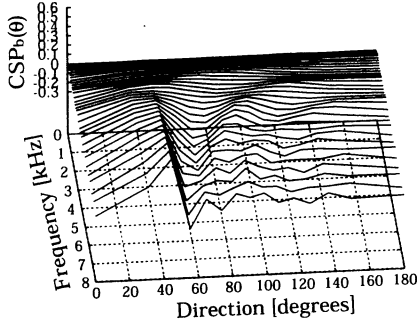


図6 各周波数帯域で得られた CSP 係数の一例

上では低い周波数帯域が細かく分割され、高い周波数帯域は荒く分割される。

その後、式(6)によって、各周波数帯域ごとに CSP 係数を計算する。

$$\text{CSP}_b(k) = \text{IDFT} \left[ \frac{s_l(\omega) \cdot s_r^*(\omega)}{|s_l(\omega)| \cdot |s_r(\omega)|} \right], \quad (6)$$

$$\omega = \omega_{L_b}, \dots, \omega_{H_b},$$

$$b = 1, \dots, B,$$

ここで、 $\text{CSP}_b(k)$  は  $b$  番目の周波数帯域で得られた CSP 係数、 $s_{[l]}(\omega)$  は  $s_{[l]}(t)$  の周波数表現、 $\omega_{L_b}$  は  $b$  番目の周波数帯域の下限周波数を、 $\omega_{H_b}$  は  $b$  番目の周波数帯域の上限周波数を、 $B$  は周波数帯域分割数を表す。図6に各周波数帯域で得られた CSP 係数の一例を示す。

### 3.2 分析重み係数

本稿では、式(7)で計算した音声の平均スペクトルに基づいて、分析重み係数を式(8)によって求めた。

$$\bar{s}(\omega) = \frac{\sum_{l=1}^L \sum_{n=1}^{N_l} |s(\omega; [l, n])|}{\sum_{l=1}^L N_l}, \quad (7)$$

$$w_b = 20 \cdot \log_{10} \left( \frac{\sum_{\omega=\omega_{L_b}}^{\omega_{H_b}} \bar{s}(\omega)}{\omega_{H_b} - \omega_{L_b}} \right), \quad (8)$$

$$b = 1, \dots, B,$$

ここで  $L$  は発話数、 $N_l$  は  $l$  番目の発話におけるフレーム数、 $s(\omega; [l, n])$  は音声のスペクトル、 $\bar{s}(\omega)$  は音声の平均スペクトル、 $\omega_{H(b)}$  は  $b$  番目の周波数帯域の上限周波数、 $\omega_{L(b)}$  は  $b$  番目の周波数帯域の下限周波数、 $w(b)$  は  $b$  番目の周波数帯域の分析重み係数を、 $B$  は帯域分割数を表す。本稿では、女性話者14名、男性話者6名の ATR503 文を使用し、フレーム長は 32 ms (ハミング窓)、フレームシフトは 16 ms の条件で学習を行った。図7に分析重み係数を示す。

### 3.3 周波数帯域選択

音声の主要な周波数成分は、第1フォルマント、第2フォルマント、第3フォルマントなどが含まれている 330 ~ 4,000 Hz に集中している。そこで、330 ~ 4,000 Hz に含まれる周波数

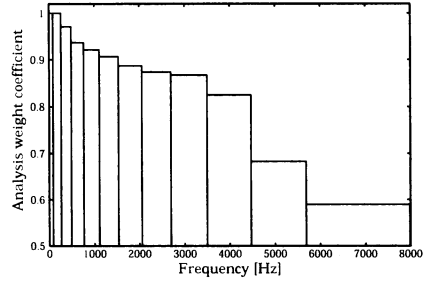


図7 分析重み係数

帯域で得られた CSP 係数を加算する。

### 3.4 話者位置推定

節3.1で計算した周波数帯域ごとの CSP 係数に、節3.2で計算した分析重み係数を乗じて加算することで、式(9)によって最終的な CSP 係数を計算する。その後、式(11)で話者方向を推定する。

$$\text{CSP}(k) = \sum_{b=1}^B w_b \cdot \text{CSP}_b(k), \quad (9)$$

$$\tau = \text{argmax}_k (\text{CSP}(k)), \quad (10)$$

$$\theta = \cos^{-1} \left( \frac{c \cdot \tau}{d \cdot F_s} \right), \quad (11)$$

ここで  $\theta$  は推定された話者方向を表す。

## 4. 評価実験

本稿では、図8に示す部屋において評価実験を行った。目的の音声はマイクロホンアレイに対して正面方向 (90°) から、雑音は右方向 (40°) 方向から到来する。マイクロホンアレイと各音源の距離は 1.5 m である。この環境において、信号対雑音比 (SNR : Signal to Noise Ratio) を変化させた場合の話者位置推定性能を評価した。さらに、提案手法によって話者方向を推定した後、マイクロホンアレイによって目的音声を強調した場合の音声認識性能も合わせて評価した。

### 4.1 実験条件

表1に示すデータ収録条件と表2に示す実験条件により評価実験を行った。データ収録を行った部屋には計算用サーバが4台置かれているため、室内騒音レベルが 50.1 dBA と非常に高

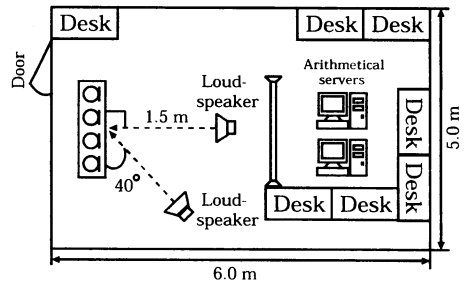


図8 実験環境

表 1 収録条件

室内騒音	50.1 dB
室内残響 T <sub>60</sub>	0.47 sec.
サンプリング周波数	16 kHz
マイクロホンアレー	8 素子, 素子感覚 21.25 mm
SNR (信号対雑音比)	-5 dB, ~, 20 dB

表 2 実験条件

テストデータ	
音声 (オープン)	ATR 音素バランス 216 単語 (女性話者 1 名, 男性話者 1 名)
雑音	白色ガウス雑音, ピンク雑音
話者位置推定	
周波数帯域分割	12
フレーム長	64 msec. (Hanning 窓)
マイクロホン間距離	148.75 (= 21.25 × 7) mm.
許容誤差	±10°
音声認識	
音響 HMM	IPA 音響モデル [13]
特徴ベクトル	MFCC, ΔMFCC, Δパワー

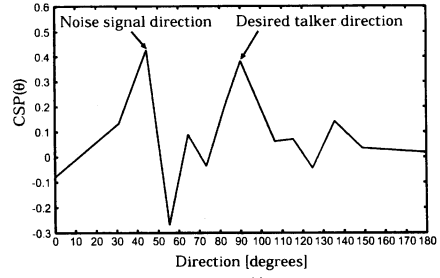
くなっている。テストデータには ATR 音素バランス 216 単語を用いた。雑音源には、全周波数で均一なスペクトル特性を持つ白色雑音と、 $-3 \text{ dB/oct.}$  のスペクトル特性を持ち、白色雑音より人間の音声の特性に近いピンク雑音を使用した。この実験条件下において、CSP 法、音声帯域 (300 ~ 4,000 Hz) を使用した CSP 法および提案手法の話者位置推定性能を評価した。話者位置推定は、マイクロホンアレーの最外郭のマイクロホンペアを使用して行った。話者位置推定性能は式 (12) に示す話者位置同定率 (TLR : Talker Localization Rate) によって評価した。

$$\text{TLR} = \frac{\sum_{l=1}^L \sum_{n=1}^{N_l} I_{\text{cor}}[l, n]}{\sum_{l=1}^L N_l} \quad (12)$$

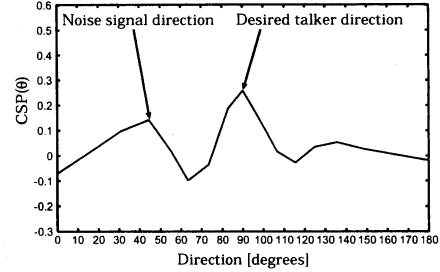
$$I_{\text{cor}}[l, n] = \begin{cases} 1 & |D_{\text{ans}}[l, n] - D_{\text{est}}[l, n]| \leq \text{Err.} \\ 0 & |D_{\text{ans}}[l, n] - D_{\text{est}}[l, n]| > \text{Err.} \end{cases}$$

ここで、TLR は話者位置同定率、 $L$  は単語数、 $N_l$  は  $l$  番目の単語におけるフレーム数、 $D_{\text{ans}}[l, n]$  は各フレームにおける正解話者位置、 $D_{\text{est}}[n]$  は各フレームにおける推定話者位置、 $\text{Err.}$  は許容誤差を表す。本稿では、 $D_{\text{ans}}[l, n]$  は常に  $90^\circ$  であり、TLR における許容誤差は  $\pm 10^\circ$  とした。また、話者位置推定は音声区間のみにおいて、64 ms のフレームごとに行った。音声区間は、収録した音声 (雑音を加算する前の) データのうち、暗騒音レベルを考慮した閾値以上のパワーを持つフレームを音声区間として切り出した。

さらに、各手法によって話者位置を推定した後、マイクロホンアレーによって目的音声を強調した場合の音声認識性能も合わせて評価した。音声認識性能は単語認識率 (WRR : Word Recognition Rate) によって評価した。



(a) CSP 法



(b) 提案手法

図 9 SNR 10 dB の白色雑音環境下における CSP 係数

## 4.2 実験結果

図 9(a)(b) に SNR 10 dB の白色雑音環境下で得られた CSP 係数を示す。ただし CSP 係数を式 (4) により時間軸  $\text{CSP}(k)$  から方向軸  $\text{CSP}(\theta)$  に変換している。図 9(a) より、CSP 法で得られる CSP 係数は、話者方向が  $90^\circ$  であるにも関わらず雑音方向の  $40^\circ$  の CSP 係数の方が大きな値になっているため、CSP 法は話者位置の推定に失敗していることが分かる。しかし、図 9(b) より、提案手法で得られる CSP 係数は、話者方向の CSP 係数が強調され雑音方向の CSP 係数よりも大きな値になっているため、提案手法は正確に話者位置を推定できることが分かる。

次に、図 10(a)(b) に話者位置推定と音声認識の実験結果を示す。図 10(a) は白色雑音環境下における実験結果を、図 10(b) はピンク雑音環境下における実験結果を示す。図 10(a)(b) 中の折れ線グラフは話者位置推定性能を、棒グラフは音声認識性能を表す。また、図 10(a)(b) の “Observed signal” はマイクロホンアレーで受音した信号から 1 ch の信号を切り出して使用した場合の実験結果、“Conventional CSP” は CSP 法の実験結果、“Conventional CSP<sub>sp</sub>” は音声帯域を使用した CSP 法の実験結果、“Subband CSP” は帯域分割型 CSP 法の実験結果を表す。

まず最初に話者位置推定実験の結果について説明する。図 10(a) より、受音信号の全帯域を使用した CSP 法よりも、音声帯域を使用した CSP 法の話者位置推定性能の方が改善されていることが分かる。これは、音声帯域を使用することで話者方向の CSP 係数が強調されたためである。さらに、提案手法の話者位置推定性能が最も高いことが分かる。次に、図 10(a) と図 10(b) を比較した場合、ピンク雑音環境下での話者位置推定性能が白色雑音環境下での話者位置推定性能よりも全体的に

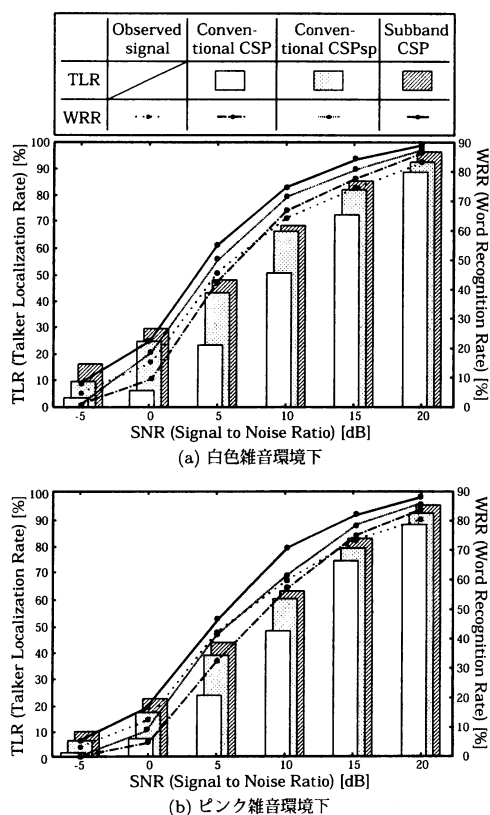


図 10 話者位置推定と音声認識の実験結果

劣化していることがわかる。これは、白色ガウス雑音が全周波数帯域で平坦なスペクトル特性を持つのにに対し、ピンク雑音は  $-3 \text{ dB/oct}$  の傾斜を持つより音声のスペクトル特性に近い雑音であるためだと考えられる。しかし、より音声に近い特性のピンク雑音環境下においても、提案手法で最も高い話者位置推定性能を得られることを確認した。

次に音声認識実験の結果について説明する。図 10(a)(b)より、CSP 法、音声帯域を使用した CSP 法で話者位置を推定した後、遅延和アレーによって音声強調を行った場合の音声認識率が、音声強調を行わない場合の音声認識率を下回っていることがわかる。これは、雑音方向にマイクロホンアレーの指向性が制御された結果、雑音が強調され、音声が悪化したためであると考えられる。図 10(a)(b)より、SNR によらず提案手法で話者位置を推定した後、遅延和アレーによって音声強調を行った場合の音声認識率が最も高くなっていることを確認した。

以上の実環境評価実験の結果より、提案した帯域分割型 CSP 法に基づく話者位置推定の有効性を確認することができた。しかし、子音部分で話者位置推定に失敗する傾向がある。これは、子音部分などのパワーが弱いフレームでは相対的に SNR が低下してしまうため、提案手法による CSP 係数強調効果が低下してしまうためであると考えられる。よって、今後はこの問題

に対処し、提案手法の話者位置推定性能をさらに改善する必要がある。

## 5. まとめ

本稿では、マイクロホンアレーを用いて遠隔発話を高品質に受信する場合に必要な発話者位置推定を正確に行うために、音声の平均スペクトルで重み付けされた帯域分割型 CSP 法に基づく話者位置推定法を提案した。暗騒音レベル 50.1 dBA の室内における実環境実験の結果、提案手法は白色雑音、ピンク雑音が存在する環境において、従来手法より高い話者位置推定性能を得られることを確認した。また、提案手法によって話者位置を推定した後にマイクロホンアレーによって目的音声を強調した場合に、音声認識率が最も改善されることを確認した。今後の課題として、目的音声の到来方向と雑音の到来方向の組み合わせを変えて提案手法の性能を評価すること、評価実験に使用する雑音の種類を増やすことなどが挙げられる。

## 謝 辞

本研究の一部は、文部科学省リーディングプロジェクト “e-Society” と科学研究費補助金 No. 14780288 による。

## 文 献

- [1] J. L. Flanagan, J. D. Johnston, R. Zahn and G. W. Elko, “Computer-Steered Microphone Arrays for Sound Transduction in Large Rooms,” *J. Acoust. Soc. Am.*, Vol. 78, No. 5, pp. 1508–1518, Nov. 1985.
- [2] L. J. Griffiths and C. W. Jim, “An Alternative Approach to Linearly Constrained Adaptive Beam-forming,” *IEEE Trans. AP*, Vol. AP-30, No. 1, pp. 27–34, 1982.
- [3] Y. Kaneda and J. Ohga, “Adaptive Microphone-array System for Noise Reduction,” *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-34, No.6, pp. 1391–1400, Dec. 1986.
- [4] 中村 哲, “音声認識系へのマイクロホンアレーの応用,” *日本音響学会講演論文集*, Vol. I, pp. 515–518, 1998.
- [5] S. U. Pillai, “Array Signal Processing,” Springer-Verlag, New York, 1989.
- [6] 大賀寿朗, 山崎芳男, 金田 豊, “音響システムとデジタル信号処理” コロナ社, 1995.
- [7] C. H. Knapp and G. C. Carter, “The generalized correlation method for estimation of time delay,” *IEEE Trans. Acoust., Speech & Signal Process.*, vol. ASSP-24, no. 4, pp. 320–327, Aug. 1976.
- [8] M. Omologo and P. Svazier, “Acoustic Event Localization in Noisy and Reverberant Environment Using CSP Analysis,” *Proc. ICASSP96*, pp. 921–924, 1996
- [9] 西浦敬信, 山田武志, 中村 哲, 鹿野 清宏, “マイクロホンアレーを用いた CSP 法に基づく複数音源位置推定,” *電子情報通信学会論文誌*, Vol. J83-D-II, No. 8, pp. 1713–1721, Aug. 2000.
- [10] I. Yamada and N. Hayashi, “Improvement of the performance of cross correlation method for identifying aircraft noise with pre-whitening of signals,” *J. Acoust. Soc. Jpn. (E)*, , vol.13, no.4, pp.241–252, 1992.
- [11] H. K. Dunn and S. D. White, “Statistical measurements on conversational speech,” *J. Acoust. Soc. Am.*, vol.11, no.3, pp.449–476, 1956.
- [12] K. Takeda, et. al., “Acoustic-phonetic labels in a Japanese speech database,” *Proc. EUROSPPECH87*, vol.2, pp.13–16, Oct. 1987.
- [13] T. Kawahara, et. al., “Japanese dictation toolkit,” *J. Acoust. Soc. Jpn. (E)*, vol.20, no.3, pp.233–239, 1999.