社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2003－SLP－47 (17)
2003／7／19

実環境下におけるマイクロホンアレーとFourier / wavelet スペクトルサブトラクションを用いた音声強調の検討

傳田 遊亀　西浦 敬信　河原 英紀

和歌山大学システム工学研究科 システム工学専攻

〒640-8510 和歌山県和歌山市栄谷930

E-Mail:{s045064, nishiura, kawahara}@sys.wakayama-u.ac.jp

あらまし　テレビ会議システムや音声制御システムにおいて，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音することは極めて重要である．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音質に受音する方法として，マイクロホンアレーやFouirerスペクトルサブトラクションによる音声強調法が提案されている．更に高品質に音声を受音することを目的として，マイクロホンアレーとFouirerスペクトルサブトラクションを組み合わせた音声強調法も提案されている．しかしこの方法では，定常雑音は効果的に抑圧できるが非定常雑音を効果的に抑圧するのは困難であるという問題があった．そこで本稿では，マイクロホンアレーとFourier / waveletスペクトルサブトラクションを組み合わせた新しい音声強調法を提案する．wavelet変換は各周波数帯域ごとに異なった時間-周波数解像度を実現できるため，waveletスペクトルサブトラクションによって非定常雑音を効果的に抑圧できると考えられる．実環境における評価実験の結果，提案手法は，従来法より高い音声認識性能と雑音抑圧性能を得られることを確認した．

キーワード　マイクロホンアレー，スペクトルサブトラクション，wavelet変換，音声認識

A Study of Speech Enhancement With Microphone Array and Fourier / wavelet Spectral Subtraction in Real Noisy Environments

Yuki DENDA　Takanobu NISHIURA　Hideki KAWAHARA

Graduate School of Systems Engineering, Wakayama University

930 Sakaedani, Wakayama, Wakayama, 640-8510 JAPAN

E-Mail:{s045064, nishiura, kawahara}@sys.wakayama-u.ac.jp

Abstract　It is very important to capture distant-talking speech with high quality for teleconferencing systems or voice-controlled systems. For this purpose, microphone array steering and Fourier spectral subtraction, for example, are ideal candidates. A combination technique using both microphone array steering and Fourier spectral subtraction has also been proposed to improve performance. However, it is difficult for the conventional approach to reduce non-stationary noise, although it is easy to robustly reduce stationary noise. To cope with this problem, we propose a new combination technique with microphone array steering and Fourier / wavelet spectral subtraction. Wavelet spectral subtraction promises to effectively reduce non-stationary noise, because the wavelet transform admits a variable time-frequency resolution on each frequency band. As a result of an evaluation experiment in a real room, we confirmed that the proposed combination technique provides better performance of the ASR (Automatic Speech Recognition) and NR (Noise Reduction) than the conventional combination technique.

Key words　Microphone array, Spectral subtraction, Wavelet transform, 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が歪みを受け，音質が低下するという問題がある．そこで，発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の研究が盛んに行われている [1]．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特性を形成することにより，高音質な音声の受音を実現している．

マイクロホンアレーの指向特性を制御する代表的な手法に遅延和アレー [2] がある．遅延和アレーは話者の方向に鋭い指向特性を形成できるため，話者方向以外から到来する雑音を抑圧できる．また，雑音の性質に依存しない効果的な抑圧が可能である．しかし，遅延和アレーの指向特性に周波数依存性があり，低周波数帯域ほど鋭い指向特性を形成できないため，遅延和アレー単体による雑音抑圧では背景雑音や室内残響を完全に抑圧することは困難である．

一方，加法性雑音（例えば空調音など）に対する抑圧手法として Fourier スペクトルサブトラクション [3] が提案されている．Fourier スペクトルサブトラクションは，観測信号のスペクトルから雑音の長時間平均スペクトルを Fourier 空間上で減算することによって加法性雑音を抑圧する．しかし，非定常雑音（例えば突発雑音）などのスペクトルと雑音の長時間平均スペクトルの間には差異が生じるため，Fourier スペクトルサブトラクションでは非定常雑音を抑圧することは困難である．

また近年，マイクロホンアレーと Fourier スペクトルサブトラクションを組み合わせた手法が提案されている [4] が，十分な非定常雑音抑圧性能が得られていない．そこで本稿では，マイクロホンアレーと Fourier / wavelet スペクトルサブトラクションを組み合わせた手法を提案し，実環境下において音声認識率を改善することを検討する．wavelet 変換は各周波数帯域ごとに異なった時間-周波数解像度を実現できるため，観測信号のスペクトルから雑音のスペクトルを wavelet 空間上で減算した場合，非定常雑音抑圧性能が改善されると考えられる．本稿ではこの手法を wavelet スペクトルサブトラクションとよぶ．

2 提案手法

本稿では，遠方で発話された音声，定常雑音と非定常雑音が同時にマイクロホンアレーに到来している状況を仮定する．このような状況で高品質な音声の受音を実現する場合，定常雑音と非定常雑音を抑圧することが必要不可欠である．本稿では，マイクロホンアレーと Fourier スペクトルサブトラクション（Fourier Spectral Subtraction:SS）と

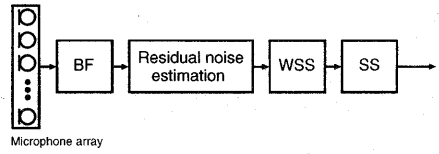


図 1: 提案手法の概要

wavelet スペクトルサブトラクション（Wavelet Spectral Subtraction:WSS）を組み合わせた手法を提案し，定常雑音と非定常雑音を抑圧することを検討する．図 1 に提案手法の概要を示す．

提案手法では，マイクロホンアレーで受音した信号に遅延和アレーによるビームフォーミング処理 (BF) を行い雑音を抑圧する．しかし，遅延和アレー単体では十分な雑音抑圧性能が得られず，出力信号に雑音が残存する場合がある．そこで，残余雑音の定常 / 非定常性に基づき，Fourier / wavelet スペクトルサブトラクションを行い残余雑音を抑圧する．Fourier スペクトルサブトラクションは定常雑音抑圧に適した手法である．一方，wavelet 変換は各周波数帯域において異なった時間-周波数解像度を実現できるため，観測信号のスペクトルから雑音のスペクトルを wavelet 空間上で減算する wavelet スペクトルサブトラクションによって，非定常雑音抑圧性能が改善されると考えられる．また，Fourier / wavelet スペクトルサブトラクションを行う前に，減算形アレー [5] に基づいた手法によって残余雑音の特性を推定する．

2.1 遅延和アレー

本稿では，遅延和アレー [2] を用いて話者の方向に鋭い指向特性を形成する．図 2 に示すように，目的の信号が θ 方向から平面波として到来し，マイクロホン数 M，マイクロホン間隔 d の等間隔直線配列マイクロホンアレーで受音される状況を考える．マイクロホンアレーで受音した信号 $x_1(t), x_2(t), \cdots, x_M(t)$ は，$x_1(t)$ に遅延を与えた信号として式 (1) として表せる．

$$x_i(t) = x_1(t - (i-1)\tau)), \tau = \frac{d\cos\theta}{c} \qquad (1)$$

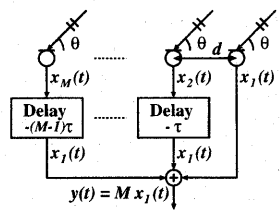


図 2: 遅延和アレー

ここで $i(i=1,2,\cdots,M)$ はチャネル番号を，c は音速を表す．遅延和アレーの出力信号 $y(t)$ は式 (2) で表せる．

$$y(t)=\sum_{i=1}^{M}x_i(t+(i-1)\tau)),\tau=\frac{d\cos\theta}{c} \tag{2}$$

式 (2) より，遅延和アレーでは θ 方向から到来する信号を同相化して加算するため，θ 方向から到来する信号は M 倍になって出力される．一方，θ 方向以外の方向から到来する信号は同相化されず M 倍にはならないため，θ 方向に感度が高く，それ以外の方向に感度が低い指向性が形成される．

2.2 Fourier スペクトルサブトラクション

2.2.1 Fourier 変換 (短時間 Fourier 変換)

周波数解析手法の 1 つである Fourier 変換を時間-周波数解析に応用したものが，短時間 Fourier 変換 [6] である．短時間 Fourier 変換は次式で定義される．

$$X(b,\omega)=\int_{-\infty}^{\infty}x(t)w(t-b)e^{-j2\pi ft}\,dt \tag{3}$$

ここで $w(t)$ は窓関数を，b は窓関数を時間軸上で移動するパラメータである．短時間 Fourier 変換では窓関数の長さは固定されており常に同じ時間-周波数解像度を持つため，定常信号の解析に適している．

2.2.2 Fourier スペクトルサブトラクション (SS)

Fourier スペクトルサブトラクション [3] は加法性雑音抑圧の効果的手法である．Fourier スペクトルサブトラクションでは，観測信号の Fourier スペクトルから雑音の長時間平均 Fourier スペクトルを減算することによって定常雑音を抑圧する．Fourier スペクトルサブトラクションは次式で定義される．

$$|\hat{X}(\omega)|=|Y(\omega)|-\alpha|\overline{N(\omega)}| \tag{4}$$

ここで $|\hat{X}(\omega)|$ は強調された音声の Fourier スペクトルを，$|Y(\omega)|$ は観測信号の Fourier スペクトルを，$|\overline{N(\omega)}|$ は雑音の長時間平均 Fourier スペクトルを，α は減算係数を表す．Fourier スペクトルサブトラクションは定常雑音抑圧には効果的であるが，非定常雑音の Fourier スペクトルと雑音の長時間平均 Fourier スペクトルの間には差異が生じるため，非定常雑音の抑圧には適していない．

2.3 wavelet スペクトルサブトラクション

2.3.1 wavelet 変換

wavelet 変換は時間-周波数解析手法の 1 つである．wavelet 変換には，各周波数帯域ごとに異なった時間-周波数解像度を実現できるという利点があるため，突発信号などの非定常信号の解析に適した手法である．wavelet 変換は次式で定義される [7]．

$$X(b,a) = \int_{-\infty}^{\infty}x(t)\overline{\psi}_{a,b}(t)\,dt \tag{5}$$

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{6}$$

ここで $\psi(t)$ はマザーウェブレットとよばれる関数である．$\overline{\psi}(\cdot)$ は $\psi(\cdot)$ の複素共役を意味し，$\psi_{a,b}(t)$ はマザーウェブレットを伸縮・平行移動することで得られる．a はスケーリングパラメータとよばれ，マザーウェブレットの伸縮の度合いを表す．b はマザーウェブレットを時間軸上で移動するパラメータである．

2.3.2 wavelet スペクトルサブトラクション (WSS)

wavelet スペクトルサブトラクションでは，観測信号の wavelet スペクトルから雑音の wavelet スペクトルを減算することによって非定常雑音を抑圧する．wavelet 変換は各周波数帯域ごとに異なった時間-周波数解像度を実現できるため，wavelet スペクトルサブトラクションは非定常雑音の抑圧に適している．wavelet スペクトルサブトラクションは次式で定義される．

$$|\hat{X}(b,a)|=|Y(b,a)|-\alpha|\overline{N(b,a)}|, \tag{7}$$

ここで $|\hat{X}(b,a)|$ は強調された音声の wavelet スペクトルを，$|Y(b,a)|$ は観測信号の wavelet スペクトルを，$|\overline{N(b,a)}|$ は雑音の wavelet スペクトルを，α は減衰係数を表す．しかし，wavelet スペクトルサブトラクションを行うためには雑音の wavelet スペクトル $|\overline{N(b,a)}|$ を正確に推定する必要がある．本稿では **2.4** において，非定常雑音の wavelet スペクトルを推定することを検討する．非定常雑音の wavelet スペクトルを正確に推定できるならば，wavelet スペクトルサブトラクションは非定常雑音を抑圧できると考えられる．

2.4 非定常雑音の wavelet スペクトル推定

wavelet スペクトルサブトラクションによって非定常雑音を抑圧するためには，非定常雑音の wavelet スペクトルを推定することが必要である．そこで，減算形アレー [5] に基づく手法によって非定常雑音の wavelet スペクトルを推定することを検討する．

2.4.1 減算形アレー

図 3 に示すように，θ 方向から平面波として到来する信号を，2 つのマイクロホン M_1, M_2 で受音する状況を考える．マイクロホン M_2 で受音した信号 $x_2(t)$ は，マイクロ

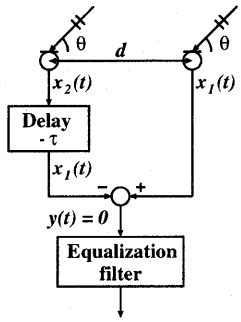


図 3: 減算形アレー

ホン M_1 で受音した信号 $x_1(t)$ に遅延を与えた信号として式 (8) として表せる．

$$x_2(t) = x_1(t-\tau), \tau = \frac{d\cos\theta}{c} \tag{8}$$

ここで，マイクロホン M_2 で受音した信号に遅延 τ を付加すれば，2 つのマイクロホンで受音した信号を同相化できる．その後，同相化した信号を減算することによって θ 方向から到来する信号を消去できる．一方，θ 方向以外から到来する信号には上記の処理によってゆがみが生じるものの，消去されることはない．

2.4.2 非定常雑音の wavelet スペクトル推定

減算形アレーを用いて観測信号から目的音を消去した信号には，定常雑音と非定常雑音がひずみを受けた状態で混在する．そこで式 (9) に示すように，減算形アレーの出力信号の wavelet スペクトルから減算形アレーの出力信号の長時間平均 wavelet スペクトルを減算することによって，非定常雑音の wavelet スペクトルをおおまかに推定する．

$$\hat{N}(b,a) = |N(b,a)| - \varepsilon(|\overline{N(b,a)}| + \sigma_{N(a)}) \tag{9}$$

ここで $\hat{N}(b,a)$ は推定された非定常雑音の wavelet スペクトルを，$|N(b,a)|$ は減算形アレーの出力信号の wavelet スペクトルを，$|\overline{N(b,a)}|$ は減算形アレーの出力信号の長時間平均 wavelet スペクトルを，$\sigma_{N(a)}$ は $|N(b,a)|$ の標準偏差を，ε は $|\overline{N(b,a)}|+\sigma_{N(a)}$ からの許容変化量を表す．式 (9) の結果，$\hat{N}(b,a) \leq 0$ なら Fourier スペクトルサブトラクションのみを行い定常雑音を抑圧する．$\hat{N}(b,a) > 0$ なら，式 (7) の $|\overline{N(b,a)}|$ の変わりに，推定された非定常雑音の wavelet スペクトル $\hat{N}(b,a)$ を用いて wavelet スペクトルサブトラクションを行い非定常雑音を抑圧する．その後，Fourier スペクトルサブトラクションを行い定常雑音を抑圧する．

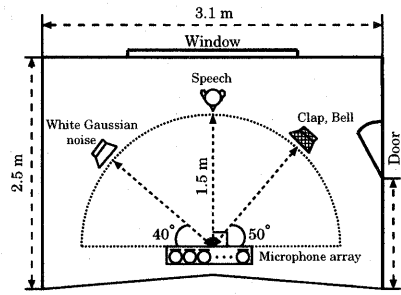


図 4: 実験環境

3 評価実験

本稿では，図 4 に示す防音室において評価実験を行い，提案手法の音声認識性能と雑音抑圧性能を評価した．目的音声はマイクロホンアレーに対して正面方向 (90 °) から，定常雑音は左方向 (40 °) から，非定常雑音は右方向 (130 °) から到来する．雑音源として，定常雑音には白色雑音を，非定常雑音には突発雑音 (拍手とベルの音 [8]) を用いた．マイクロホンアレーと各音源の距離は 1.5 m とした．この環境において，SNR を変化させたときの音声認識性能および雑音抑圧性能を評価した．

3.1 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．この実験条件下においてシングルマイクロホンおよびマイクロホンアレーを用いて，定常雑音対非定常雑音比 (stationary Noise to non-stationary Noise Ratio:NNR) が 0 dB，信号対雑音比 (Signal to Noise Ratio:SNR) が -5 dB, ∼, 20 dB, clean における音声認識性能と雑音抑圧性能を評価した．なお本稿では，目的音の到来方向は既知であるとし，マザーウェブレットには Gabor 関数 [7] を用い，分析オクターブ数を 6，オクターブ内分析数を 8 とした．テストデータには ATR 音素バランス 216 単語を用いた．音声認識性能は単語認識率 (Word Recognition Rate:WRR) によって，雑音抑圧性能は雑音抑圧率 (Noise Reduction Rate:NRR) によって評価した．雑音抑圧率は出力 $-SNR$ から 入力 $-SNR$ を減算することによって求めた．

3.2 予備評価実験

予備実験として，定常雑音のみと非定常雑音のみの環境における性能を評価した．図 5(a) は定常雑音のみの環境における実験結果を表す．図 5(b) は非定常雑音のみの環境における実験結果を表す．図中の折れ線グラフは単語認識率

表 1: 収録条件

マイクロホンアレー	8 素子, 2.125 cm 間隔
サンプリング周波数	16 kHz
室内残響 $T_{[60]}$	0.12 sec
室内騒音	20 dBA
SNR(信号対雑音比)	-5 dB, ～, 20 dB, clean
NNR	0 dB
(定常雑音対非定常雑音比)	

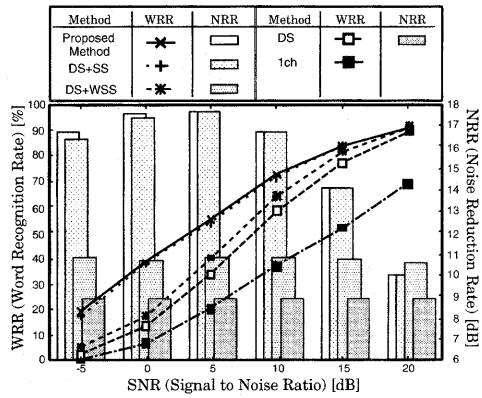
表 2: 実験条件

フレーム長	32 msec (Hanning 窓)
フレーム周期	8 msec
減衰係数 (α)	SS: 1.0　WSS: 1.0
許容変化量 (ε)	WSS: 2.0
音声認識	
音響 HMM	IPA 音響モデル [9]
特徴ベクトル	MFCC, ΔMFCC, Δ パワー
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, 男性話者 1)
定常雑音	白色雑音
非定常雑音	拍手, ベル (RWCP データベース [8])

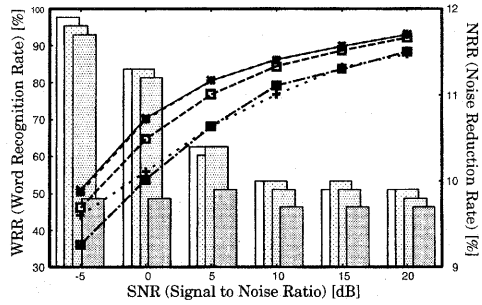
を，棒グラフは雑音抑圧率を示す．図中の 1 ch はシングルマイクロホンで音を受音した場合の単語認識率と雑音抑圧率の結果を表す．

最初に図 5(a) の目的音声と定常雑音の環境における結果について説明する．図の折れ線グラフより，提案手法，遅延和アレーと Fourier スペクトルサブトラクションを組み合わせた場合，遅延和アレーと wavelet スペクトルサブトラクションを組み合わせた場合のすべてにおいて，雑音を抑圧しない場合の単語認識率を上回る結果を得られることが確認でき，その中でも提案手法の単語認識率が一番高いことが確認できた．さらに定常雑音のみの環境においては，遅延和アレーと Fourier スペクトルサブトラクションを組み合わせた場合のほうが，遅延和アレーと wavelet スペクトルサブトラクションを組み合わせた場合よりも効果的であることがわかり，Fourier スペクトルサブトラクションが定常雑音の抑圧に効果的であることが確認できた．

次に図 5(b) の目的音声と非定常雑音の環境における結果について説明する．図の折れ線グラフより，提案手法，遅延和アレーと Fourier スペクトルサブトラクションを組み合わせた場合，遅延和アレーと wavelet スペクトルサブトラクションを組み合わせた場合のすべてにおいて，雑音を抑圧しない場合の単語認識率を得られることが確認でき，その中でも提案手法の単語認識率が一番高いことが確認できた．さらに非定常雑音のみの環境においては，遅延和アレーと wavelet スペクトルサブトラクションを組み合わせた場合のほうが，遅延和アレーと Fourier スペクトルサブトラクションを組み合わせた場合よりも効果的であることがわかり，wavelet スペクトルサブトラクションが非定常雑音の抑圧に効果的であることが確認できた．



(a) 目的音声と定常雑音が混在する環境



(b) 目的音声と非定常雑音が混在する環境

図 5: 単語認識率と雑音抑圧率

3.3 目的音声，定常雑音と非定常雑音が混在する環境における評価実験

最後に，目的音声，定常雑音と非定常雑音が混在する環境において提案手法の性能を評価した．図 6(a) は目的音声，図 6(b) は定常雑音，図 6(c) は非定常雑音，図 6(d) は受音信号 (目的音声+定常雑音+非定常雑音)，図 6(e) は遅延和アレーによって雑音を抑圧した音声，図 6(f) は遅延和アレー+Fourier スペクトルサブトラクションによって雑音を抑圧した音声，図 6(g) は遅延和アレー+wavelet スペクトルサブトラクションによって非定常雑音を抑圧した音声，図 6(h) は提案手法によって雑音を抑圧した音声の波形をそれぞれ示している．

図 6(e) より，遅延和アレー単体では定常雑音，非定常雑音ともに十分に抑圧されていないことがわかる．次に図 6(f) より，遅延和アレーと Fourier スペクトルサブトラクションを組み合わせることで定常雑音をほぼ抑圧できていることがわかる．しかし，非定常雑音は Fourier スペクトルサブトラクションでは抑圧されていないことがわかる．次に図 6(g) より，遅延和アレーと wavelet スペクトルサブトラクションを組み合わせることで非定常雑音を抑圧できていることがわかる．最後に図 6(h) より，提案手法は定常雑音，非定常雑音ともに抑圧できていることがわかる．

次に，図 7 に音声認識性能と雑音抑圧性能の結果を示す．図中の折れ線グラフは単語認識率を，棒グラフは雑音抑圧率を示す．図中の 1 ch はシングルマイクロホンで音を受音した場合の単語認識率と雑音抑圧率の結果を表す．図の折れ線グラフより提案手法は，遅延和アレーと Fourier スペクトルサブトラクションを組み合わせた場合，遅延和アレーと wavelet スペクトルサブトラクションを組み合わせた場合，雑音を抑圧しない場合のすべてを上回る単語認識率を得られることが確認できた．

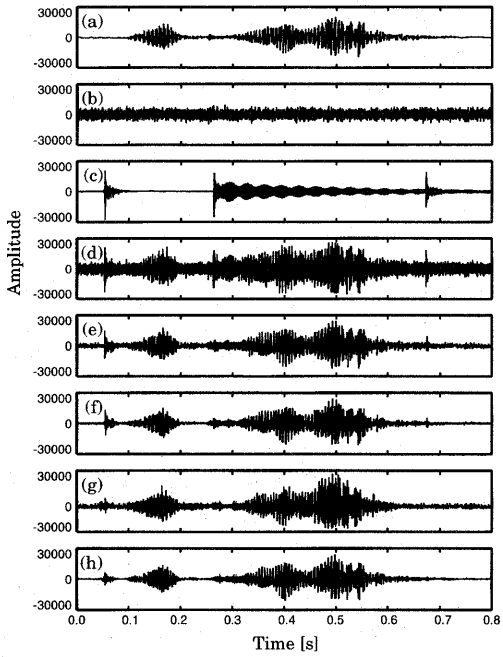


図 6: 波形の一例:(a) 目的音声，(b) 定常雑音，(c) 非定常雑音，(d) 受音信号 (目的音声+定常雑音+非定常雑音)，(e) 雑音抑圧後の音声 (遅延和アレー)，(f) 雑音抑圧後の音声 (遅延和アレー+Fourier スペクトルサブトラクション)，(g) 雑音抑圧後の音声 (遅延和アレー+wavelet スペクトルサブトラクション)，(h) 雑音抑圧後の音声 (提案手法)

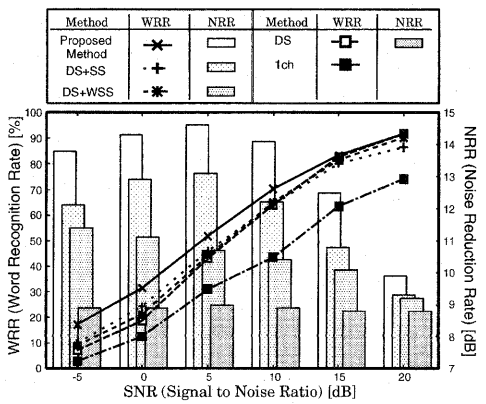


図 7: 目的音声，定常雑音と非定常雑音が混在する環境における単語認識率と雑音抑圧率

4 まとめ

本稿では，遠方で発話された音声を高品質に受音するために，マイクロホンアレーと Fourier / wavelet スペクトルサブトラクションを組み合わせることを検討した．その結果，定常雑音のみの環境下，非定常雑音のみの環境下と定常雑音と非定常雑音が混在する環境下において，従来法よりも音声認識性能と雑音抑圧性能を改善できることを確認した．今後の課題として，非定常雑音の wavelet スペクトル推定精度を向上させ非定常雑音抑圧性能をより改善すること，また，指向性雑音環境下だけでなく拡散性雑音環境下で提案手法の性能評価実験を行うことが挙げられる．

参考文献

[1] 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.

[2] 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.

[3] S. F. Boll, “Suppression of Acoustic Noise in Speech Using Spectral Subtraction,” IEEE Trans. ASSP, Vol. ASSP-27, No. 2, pp. 113–120, Apr. 1979.

[4] M. Dahl, et. al., “Simultaneous Echo Cancellation and Car Noise Suppression Employing a Microphone Array,” ICASSP97, Vol.1, pp. 239–242, Apr. 1997.

[5] 金田 豊, “マイクロホンアレーによる指向性制御,” 日本音響学会誌, vol. 51, no. 5, pp. 390–394, 1995.

[6] D. Gabor, “Theory of Communications,” J. IEE, Vol. 93, No. 26, pp. 429–457, Nov. 1946.

[7] I. Daubechies, “The Wavelet Transform, Time-frequency Localization and Signal Analysis,” IEEE Trans. Inf. Theory, Vol 36, pp. 961–1005, Sep. 1990.

[8] S. Nakamura, et. al., “Data Collection in Real Acoustical Environments for Sound Scene Understanding and Hands-Free Speech Recognition,” Proc. Eurospeech99, pp. 2255–2258, Sep. 1999.

[9] T. Kawahara, et. al., “Japanese Dictation Toolkit,” J. Acoust. Soc. Jpn. (E), Vol. 20, No. 3, pp. 233–239, 1999.