音声言語情報処理 42−12
(2002. 7. 13)

雑音除去とモデル適応を併用した雑音下音声認識 — AURORA2タスクでの評価 —

藤本 雅清　　有木 康雄

龍谷大学 理工学部
〒 520-2194 大津市瀬田大江町横谷 1-5　　Tel: 077-543-7427
E-mail: masa@arikilab.elec.ryukoku.ac.jp, ariki@rins.st.ryukoku.ac.jp

あらまし 本研究では，雑音除去法と音響モデル適応法を併用した，雑音に頑健な音声認識法を提案し，AURORA2タスクでの評価を行った．雑音除去手法には二つの方法を用いており，一つは短時間フレーム及び周波数帯域ごとに雑音スペクトルの減算量を変化させる，帯域分割型適応スペクトルサブトラクション (ASBSS) 法であり，もう一方はASBSS法により得られた音声スペクトルをカルマンフィルタにより再推定する方法である．本研究では，これら二つの方法を併用することにより，精度良く音声スペクトルを推定することについて検討を行った．また，一般に雑音除去を行うと，推定誤差等による残差雑音が生じてしまい，音声認識率に影響を与えるという問題がある．この問題を解決するために，本研究では教師無しMLLR適応を用いることにより，残差雑音により生じるスペクトル歪みに音響モデルを適応させた．本手法をAURORA2データベースを用いて評価した結果，Clean Training Condition，Multi Training Conditionともに大幅な認識率の改善が得られた．

キーワード ：雑音下での音声認識，雑音除去，音響モデル適応，AURORA2データベース

Noisy Speech Recognition Based on Noise Reduction and Acoustic Model Adaptation — An Evaluation on the AURORA2 Tasks —

Masakiyo Fujimoto　　Yasuo Ariki

Faculty of Science and Technology, Ryukoku University
1-5 Yokotani, Oe-cho, Seta, Otsu-shi, 520-2194 Japan　　Tel: +81-77-543-7427
E-mail: masa@arikilab.elec.ryukoku.ac.jp, ariki@rins.st.ryukoku.ac.jp

Abstract In this paper, we have evaluated a noisy speech recognition method based on noise reduction and acoustic model adaptation, on the AURORA2 tasks. For noise reduction method, we employed two noise reduction methods. One is an Adaptive Sub-Band Spectral Subtraction (ASBSS) method which can optimize the noise subtraction rate according to the SNR in frequency bands at each frame. The other is a Kalman filtering estimation method which re-estimates the accurate speech spectra from those estimated by ASBSS. The accurate speech spectra was estimated by combining these two methods. Usually, a noise reduction method has a problem that it degrades the recognition rate because of spectral distortion caused by residual noise occurred through noise reduction and over estimation. To solve the problem in the noise reduction method, adaptation of the acoustic models is employed by using an unsupervised MLLR adaptation to the spectral distortion. In evaluation on the AURORA2 tasks, our method showed the significant improvement in recognition accuracy for both clean training condition and multi training condition.

Key words : noisy speech recognition, noise reduction, acoustic model adaptation, AURORA2 database

1 はじめに

ここ数年，数多くの音声認識手法が提案されており，また，音声認識システムを実装したソフトウェア，家電製品等が商品化され，音声認識の実用化が進められている．しかし，それらの多くは比較的静かな環境を想定したものが大半を占めており，実環境で背景雑音の影響が大きい場合，認識精度が極端に低下してしまうという問題があり，完全な実用化には至っていないのが現状である．これを受けて，背景雑音に頑健な音声認識システムを確立し，音声認識システムの実用化を実現するために，様々な研究が行われている[1, 2, 3].

雑音に頑健な音声認識システム確立のためのアプローチとして，認識システムを雑音に適応させる方法(雑音適応)[4]-[6]と，雑音が重畳した音声から雑音成分を取り除き，クリーンな音声を抽出して認識を行う方法(雑音除去)[7]-[10]の2種類が考えられる．

雑音除去の方法として従来，Spectral Subtraction(SS)法[7]がよく用いられており，少ない計算量の割には有効な手法である．SS法は雑音が重畳した音声のパワースペクトルから，雑音の推定パワースペクトルを減算することにより，音声のパワースペクトルを推定する手法であり，SNRに応じて雑音パワースペクトルを減算する割合を変化させることにより，高い推定精度が得られることが知られている．

一般的なSS法において，雑音パワースペクトルを減算する割合は入力信号の全フレームにおいて一律の値が設定されていることが多い．しかし，雑音が比較的定常であっても，1フレーム単位で見た，局所的なSNR(Segmental SNR)はクリーン音声のパワーに依存して常に変化しており，このSegmental SNRに応じて各フレームにおける雑音パワースペクトルの減算量を変化させることによって，高い音声パワースペクトルの推定精度が得られると考えられる．この方法は，適応Spectral Subtraction(Adaptive Spectral Subtraction: ASS)法と呼ばれており，山本らにより提案されている[8].

また，母音と子音のパワースペクトル特徴の違いを考慮すると，全周波数帯域で一様に減算するのでは無く，周波数帯域に応じても雑音パワースペクトルの減算量を変化させることにより，より高い音声パワースペクトルの推定精度が得られると考えられる．この方法は，非線型Spectral Subtraction(Non-linear Spectral Subtraction: NSS)法と呼ばれており，Lockwoodらにより提案されている[9].

以上のことをふまえた上で，本研究ではASS法とNSS法を組み合わせることにより，短時間フレーム及び周波数帯域ごとに雑音スペクトルの減算量を変化させる，帯域分割型適応スペクトルサブトラクション(ASBSS)法[10]を用いる．また，より高い音声パワースペクトルの推定精度を得るために，カルマンフィルタを用いて，ASBSS処理後の音声パワースペクトルを再推定することについても検討した．

ここで，一般に雑音除去を行うと，推定誤差等による残差雑音及び，音声パワースペクトルの歪みが生じてしまい，音声認識率に影響を与えるという問題がある．この問題を解決するために，本研究では教師無しMLLR適応[11]を用いることにより，推定誤差により生じるスペクトル歪みに音響モデルを適応させた．

本手法の評価には，AURORA2と呼ばれる雑音下音声認識の評価用データベースを用いて行っている．AURORA2データベースを用いて評価を行った結果，AURORA2データベースに含まれる全ての雑音環境において，大幅な認識率の改善が得られた．

2 AURORA2データベース

本研究で使用したAURORA2データベースの詳細について述べる．AURORA2データベースは仏国ELRA(European Language Resources Association)[12]より配布されている，雑音下音声認識の評価用データベースである．AURORA2データベースに含まれる雑音重畳音声データは，米国LDC(Linguistic Data Consortium)[13]より配布されているTI-DIGITS(連続英語数字音声データベース)データベースに種々の雑音を計算機上で重畳することにより生成されており，表1に示すような，3種類のテストセットが用意されている[14].

表 1: AURORA2データベースに含まれる雑音環境

テストセット名	雑音環境	フィルタ特性
SetA	Subway, Babble Car, Exhibition	G.712
SetB	Restaurant, Street Airport, Station	G.712
SetC	Subway, Street	MIRS

表1において，SetA，SetBではそれぞれ4種類，SetCではSetA，SetBから1種類ずつ選択した雑音が用いられ，SNRは-5～20dB(5dB刻み)及びクリーン環境が用意されている．全ての音声データには，電話回線を模擬したフィルタ特性が畳み込まれており，SetA，SetBではG.712，SetCではMIRSと呼ばれるフィルタ特性になっている[14]．また，各雑音，SNRごとに1001文章の音声データ(男女混在)が含まれており，各音声データの標本化周波数は8kHz(16Bit)である．

次に認識システムと，評価方法について述べる．HMMの学習及び認識は，HTK(Hidden Markov Model Toolkit)[15]により行われており，学習，認識を行うためのスクリプトが提供されている．認識時の語彙

数は13(数字1～9, oh, zero, 無音，ショートポーズ)であり，各語彙ごとにHMMを学習する．AURORA2データベース標準のHMMの構造は表2の通りであり，ショートポーズのHMMは無音HMMの第3状態を共有している．また，全てのHMMにおいて状態のスキップは考慮されていない．

表 2: AURORA2データベース標準HMMの構造

	状態数	混合分布数
数字(1～9, oh, zero)	18状態16ループ	3
無音	5状態3ループ	6
ショートポーズ	3状態1ループ	6

HMMの学習データセットとしては，クリーン音声のみの学習データセット(Clean Training Condition)と，雑音重畳音声を含んだ学習データセット(Multi Training Condition)の2種類の学習データセットが用意されており，それぞれの学習データセットを用いてHMMを学習する．Multi Training Conditionに含まれる雑音重畳音声データには，テストセットSetAに含まれる4種類の雑音が重畳しており，SNRはクリーン及び，0～20dBのみである．HMM学習及び認識時の音響分析の条件は表3のように定められている(ベースラインシステムではCMN処理は行わない)．

表 3: AURORA2データベースにおける音響分析条件

高域強調	$1-0.97z^{-1}$
特徴量	12次MFCC + Log-Power + 12次ΔMFCC + ΔLog-Power + 12次ΔΔMFCC + ΔΔLog-Power
分析区間長	25ms
分析周期	10ms
時間窓	Hamming Window

音声認識時には，クリーン音声により学習されたHMMと，雑音重畳音声により学習されたHMMそれぞれを用いて評価を行う．評価は単語(数字)誤り率及び，表4に示したベースラインの単語誤り率に対する改善率により行う．実質的な評価に用いられるのは，各テストセットのSNR0～20dBにおける単語誤り率，改善率の平均値であり，クリーン及び-5dB環境は評価の対象とはならない．

表 4: AURORA2データベースのベースライン認識率

Aurora 2 Reference Word Error Rate				
	Set A	Set B	Set C	Overall
Multi	11.93%	12.78%	15.44%	12.97%
Clean	41.26%	46.60%	34.00%	41.94%
Average	26.59%	29.69%	24.72%	27.46%

本研究では，以上のようなデータベースを用いて，提案手法の評価を行っている．なお，2001年開催のEurospeech，2002年開催のICSLPといった音声研究の国際会議において，このAURORA2データベースを用いたスペシャルセッションが立ち上げられている．

3 帯域分割適応スペクトルサブトラクション法

本研究で提案する帯域分割適応スペクトルサブトラクション法について述べる．

3.1 適応スペクトルサブトラクション法

雑音重畳音声のパワースペクトルを$X(f,i)$(fはFFT分析におけるチャネル番号，iはフレーム番号)，クリーン音声の推定パワースペクトルを$\hat{S}(f,i)$，雑音の平均推定パワースペクトル(数フレーム分の雑音スペクトルの平均)を$\overline{N}(f)$としたとき，SS法は以下のように示される(α, βはそれぞれサブトラクション係数，フロアリング係数を表す)．

$$\hat{S}(f,i) = \max\left[X(f,i) - \alpha\overline{N}(f), \beta X(f,i)\right] \quad (1)$$

一般に，SNRに応じてサブトラクション係数αの値を変化させることにより，高い音声パワースペクトルの推定精度が得られることが知られている．ここで，一般に言われるSNRとは，音声全体での平均値のことであり，雑音が比較的定常であっても，1フレーム単位で見た，局所的なSNR(Segmental SNR)はクリーン音声のパワーに応じて常に変化している．よって，1フレーム単位のSegmental SNR($SNR(i)$と定義する)に応じて，各フレームにおけるサブトラクション係数$\alpha(i)$の値を，式(2)のような決定関数gを定義して設定すれば，より高い音声パワースペクトルの推定精度が得られるものと考えられる．この方法は，適応Spectral Subtraction(Adaptive Spectral Subtraction: ASS)法と呼ばれており，山本らにより提案されている[8]．

$$\alpha(i) = g(SNR(i)) \quad (2)$$

次に，$SNR(i)$の推定法について述べる．雑音重畳音声の短時間RMS(Root Mean Square)パワーを$Pow_x(i)$，クリーン音声の推定短時間RMSパワーを$Pow_s(i)$，雑音の平均推定短時間RMSパワー(数フレーム分の雑音短時間RMSパワーの平均)を$\overline{Pow_n}$としたとき，$SNR(i)$は以下のように推定される．

$$SNR(i) = \begin{cases} 20\log_{10}\dfrac{Pow_s(i)}{\overline{Pow_n}} & Pow_s(i) > 0 \\ \gamma \ (=-10) & Pow_s(i) \leq 0 \end{cases} \quad (3)$$

$$Pow_s(i) = Pow_x(i) - \overline{Pow_n} \quad (4)$$

$Pow_s(i)$が負の値を持つとき，$SNR(i)$を計算できないので，定数γを代入する．

次に，得られたSegmental SNRである$SNR(i)$を用いて，サブトラクション係数$\alpha(i)$を与えるサブトラクション係数決定関数$g(SNR(i))$を構成する．本研究ではサブトラクション係数決定関数$g(SNR(i))$として図1に示すような関数を与えており，$SNR(i)$が0dB以下の場合に$\overline{N}(f)$は減算量が最大(2.0倍)となり，30dB以上の場合には減算を行っていない．尚，この関数fの形状は実験的に求めたものである．また，フロアリング係数は$\beta = 0.2$としている．

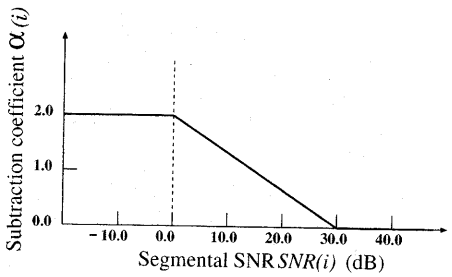

図 1: サブトラクション係数決定関数 $g(SNR(i))$

3.2 帯域分割適応スペクトルサブトラクション法への拡張

3.1のASS法では，フレームごとにサブトラクション係数の値を変化させているが，全周波数帯域で一様の割合で減算を行っている．しかし，母音（低域にエネルギーが集中）と子音（高域にエネルギーが集中）のパワースペクトル特徴の違いを考慮に入れると，全周波数帯域で一様の割合で減算するのでは無く，周波数帯域においてもサブトラクション係数の値を変化させる必要があると考えられる．

以上のことから，本研究では，短時間フレーム及び周波数帯域ごとに雑音スペクトルの減算量を変化させる，帯域分割型適応スペクトルサブトラクション(Adaptive Sub-Band Spectral Subtraction: ASBSS)法[10]を用いる．ASBSS法の処理手順は以下のようになっており，処理の概要図を図2に示す．

1. 雑音重畳音声を8チャネルのメル尺度バンドパスフィルタに通して，サブバンド波形に分解する．
2. 分解されたサブバンド波形の雑音のRMSパワーと雑音重畳音声のRMSパワーを用いて，各サブバンドにおけるSegmental SNR $SNR(k,i)$を推定する(kはサブバンドのチャネル番号)．
3. Segmental SNR $SNR(k,i)$に応じて各帯域でのサブトラクション係数$\alpha(k,i)$を決定する．
4. 得られたサブトラクション係数$\alpha(k,i)$を用いて雑音パワースペクトルを減算し，クリーン音声のパワースペクトルを推定する．

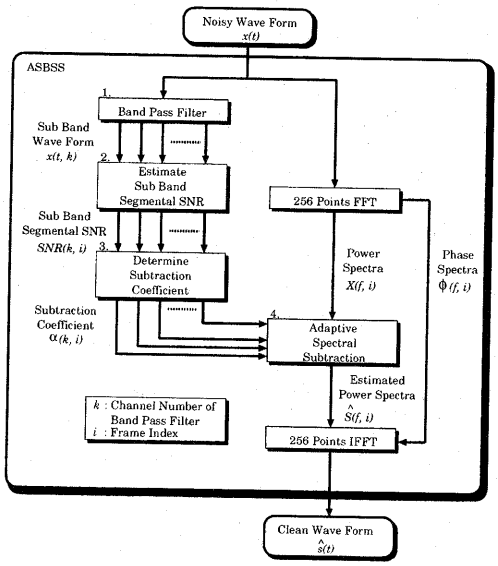

図 2: ASBSS法の処理概要図

4 カルマンフィルタを用いた再推定

カルマンフィルタに基づく音声パワースペクトルの再推定法について以下に述べる．

4.1 状態空間モデル

音声パワースペクトル$S(f,i)$をカルマンフィルタにより再推定するために，本研究では以下のような状態空間モデルを定義した．

$$\tilde{S}^l(f,i+1) = F_{f,i}\tilde{S}^l(f,i) \tag{5}$$

$$F_{f,i} = \frac{\hat{S}^l(f,i+1)}{\hat{S}^l(f,i)} \tag{6}$$

$$\begin{aligned} X^l(f,i) &= \log\left(\tilde{S}(f,i) + \tilde{N}(f,i)\right) \\ &= \tilde{S}^l(f,i) + \log\left(1 + \frac{\tilde{N}(f,i)}{\tilde{S}(f,i)}\right) \\ &= \tilde{S}^l(f,i) + V(f,i) \end{aligned} \tag{7}$$

$$\tilde{N}(f,i) = X(f,i) - \tilde{S}(f,i) \tag{8}$$

$$V(f,i) = \log\left(1 + \frac{\tilde{N}(f,i)}{\tilde{S}(f,i)}\right) \tag{9}$$

$\tilde{S}(f,i)$はカルマンフィルタによる$S(f,i)$の再推定値，$\hat{S}(f,i)$はASBSS法による$S(f,i)$の推定値，添字lは対数パワースペクトル領域を示す．また，上式において，式(5)は状態空間モデルにおける状態方程式，式(7)は観測方程式である．

4.2 カルマンフィルタによる推定

4.1の状態空間モデルにより，以下のようなカルマンフィルタのフィルタ方程式が得られる．これらのフィルタ方程式を用いて，$S(f,i)$の再推定を逐次的に行う．

$$\tilde{\mathbf{S}}^{\mathbf{l}}_i = \mathbf{F}_{i-1}\tilde{\mathbf{S}}^{\mathbf{l}}_{i-1} + \mathbf{K}_i(\mathbf{X}^{\mathbf{l}}{}_i - \mathbf{F}_{i-1}\tilde{\mathbf{S}}^{\mathbf{l}}_{i-1}) \quad (10)$$

$$\mathbf{K}_i = \mathbf{Q}_i\left[\mathbf{Q}_i + \mathbf{\Sigma}_{\mathbf{V}_i}\right]^{-1} \quad (11)$$

$$\mathbf{Q}_i = \mathbf{F}_{i-1}(\mathbf{I} - \mathbf{K}_{i-1})\mathbf{Q}_{i-1}\mathbf{F}^T_{i-1} \quad (12)$$

$$\tilde{\mathbf{S}}^{\mathbf{l}}_i = \left(\tilde{S}^l(0,i),\cdots,\tilde{S}^l(N-1,i)\right)^T \quad (13)$$

$$\mathbf{X}^{\mathbf{l}}{}_i = \left(X^l(0,i),\cdots,X^l(N-1,i)\right)^T \quad (14)$$

$$\mathbf{F}_i = diag\left(F_{0,i},\cdots,F_{N-1,i}\right) \quad (15)$$

Nは，FFT分析の次元数，$\mathbf{Q}_i$は推定誤差の共分散行列であり，カルマンフィルタの初期値はそれぞれ以下のように設定した．

$$\tilde{\mathbf{S}}^{\mathbf{l}}_0 = \left(\hat{S}^l(0,0),\cdots,\hat{S}^l(N-1,0)\right) \quad (16)$$

$$\mathbf{Q}_0 = \eta\cdot\mathbf{I} \quad (\eta = 0.001) \quad (17)$$

Eq.(11)の$\mathbf{\Sigma}_{\mathbf{V}_i}$は，$\mathbf{V}_i\left(=(V(0,i),\cdots,V(N-1,i))^T\right)$の共分散行列であり，平均零の白色雑音であると仮定することにより，以下のようにして求めている．

$$\mathbf{\Sigma}_{\mathbf{V}_i} = \mathbf{V}_i\mathbf{V}_i^T \quad (18)$$

5 教師無し MLLR適応

一般に雑音除去を行うと，推定誤差等による残差雑音及び，音声スペクトルの歪みが生じてしまい，音声認識率に影響を与えるという問題がある．この問題を解決するために，本研究では教師無しMLLR適応[11]を用いることにより，推定誤差により生じるスペクトル歪みに音響モデルを適応させた．教師無しMLLR適応を数字HMMに対して行うためには，適応データの数字ラベルが必要となる．本研究では，適応データを適応前のHMMにより認識した結果を数字ラベルとして用いている．また，適応データには入力音声1文章のみを用いており，MLLR適応におけるHMM内の正規分布クラスタ数は1とした．

6 実験

以上に述べた方法を，AURORA2データベースを用いて評価した．音声認識の評価は，2で述べたAURORA2データベース標準のHMMと，Microsoft社のAsela Gunawardanaによって提供された，高精度HMM[16, 17]の両方を用いて行っている．ここで高精度HMMでは，状態数は表2の標準HMMと同じであるが，数字HMMの混合分布数が20，無音，ショートポーズHMMの混合分布数が36という構造になっている．音響分析の条件は，表3の条件に従っており，それぞれのHMMを用いた認識において，以下の二つの方法で評価を行っている．

手法(1) ： ASBSS法 + MLLR

手法(2) ： ASBSS法 + Kalman filter + MLLR

6.1 標準HMMによる実験結果

標準HMMによる認識結果を表5，6に示す．

表 5: 手法(1)による認識結果(標準HMM)

Aurora 2 Word Error Rate				
	Set A	Set B	Set C	Overall
Multi	11.94%	13.78%	11.67%	12.62%
Clean	20.25%	21.49%	23.27%	21.35%
Average	16.09%	17.63%	17.47%	16.98%

Aurora 2 Relative Improvement				
	Set A	Set B	Set C	Overall
Multi	-3.28%	-10.40%	14.25%	2.62%
Clean	49.34%	59.76%	42.12%	52.07%
Average	23.03%	24.68%	28.18%	24.72%

表 6: 手法(2)による認識結果(標準HMM)

Aurora 2 Word Error Rate				
	Set A	Set B	Set C	Overall
Multi	10.22%	12.31%	10.96%	11.21%
Clean	27.66%	28.52%	25.76%	27.62%
Average	18.94%	20.42%	18.36%	19.41%

Aurora 2 Relative Improvement				
	Set A	Set B	Set C	Overall
Multi	8.44%	7.24%	22.43%	10.76%
Clean	30.99%	45.23%	27.43%	35.97%
Average	19.71%	26.23%	24.93%	23.36%

表5，6より，Clean Training Conditionでは手法(1)により大幅な改善が得られ，Multi Training Conditionでは手法(2)により大幅な改善が得られた．特に手法(2)のClean Training Conditionでは低SNRでの改善率が低くなっていた．手法(2)がClean Training Conditionでの改善率が低いにも関わらず，Multi Training Conditionで高い改善率を得ている理由として，以下の2点が理由として挙げられる．

(I):カルマンフィルタでは状態遷移係数$F_{f,i}$の信頼性が$S(f,i)$の推定精度に大きく依存する．また，式(6)に示すように，$F_{f,i}$はASBSS法により得られた対数パワースペクトル$\hat{S}^l(f,i+1)$と$\hat{S}^l(f,i)$から計算されるため，$F_{f,i}$の信頼性はASBSS法の精度に依存す

る．しかし，低SNR環境ではASBSS法の精度が劣化するため，$F_{f,i}$の信頼性が低下する．このことより，低SNRでは$F_{f,i}$の信頼性の劣化により，カルマンフィルタによる推定誤差が大きくなり，スペクトルの歪みが発生したと考えられる．

(II):今回の評価ではHMMの学習は，雑音除去処理が行われた学習データを用いて行っている．ここでMulti Training Conditionでは，雑音が重畳した学習データに対して雑音除去処理を施してから学習を行っているため，学習データの低SNRのスペクトル歪みがHMMに反映される．一方，Clean Training Conditionでは与えられた学習データがクリーンのみなので，低SNRのスペクトル歪みがHMMに反映されない．

これらの理由により，手法(2)がMulti Training Conditionで高い改善率を得たと考えられる．

6.2 高精度HMMによる実験結果

高精度HMMによる結果を表7，8に示す．

表 7: 手法(1)による認識結果(高精度HMM)

Aurora 2 Word Error Rate				
	Set A	Set B	Set C	Overall
Multi	8.43%	10.85%	8.93%	9.50%
Clean	15.80%	16.45%	18.61%	16.62%
Average	12.11%	13.65%	13.77%	13.06%

Aurora 2 Relative Improvement				
	Set A	Set B	Set C	Overall
Multi	36.29%	20.72%	42.92%	31.39%
Clean	66.31%	72.44%	58.07%	67.11%
Average	51.30%	46.58%	50.49%	49.25%

表 8: 手法(2)による認識結果(高精度HMM)

Aurora 2 Word Error Rate				
	Set A	Set B	Set C	Overall
Multi	7.83%	10.01%	9.49%	9.03%
Clean	25.77%	27.41%	27.53%	26.78%
Average	16.80%	18.71%	18.51%	17.91%

Aurora 2 Relative Improvement				
	Set A	Set B	Set C	Overall
Multi	40.50%	29.53%	41.72%	36.35%
Clean	43.14%	51.84%	30.18%	44.03%
Average	41.82%	40.68%	35.95%	40.19%

それぞれの表において，6.1の結果に比べて大幅な改善が得られている．また，それぞれの手法において6.1の結果と同様に，手法(2)がMulti Training Conditionにおいて改善される傾向が現れている．

7 おわりに

本研究では，雑音除去法と音響モデル適応法を併用した，雑音に頑健な音声認識法を提案し，AURORA2データベースを用いて評価を行った．本手法をAURORA2データベースを用いて評価した結果，Clean Training Condition，Multi Training Conditionともに大幅な認識率の改善が得られた．今後より高精度な雑音除去法について検討し，Clean Training Conditionでの改善を目指す予定である．

謝辞

本研究を行うにあたり多大な助言を頂いた，SLP雑音下音声認識評価ワーキンググループの皆様方に深く感謝致します．

参考文献

[1] 中川聖一: "ロバストな音声認識のための音響信号処理", 音響誌, 53巻, 11号, pp.864-871(1997).
[2] 松本 弘: "音声認識における環境適応技術", 信学技報, SP99-111, pp.109-114(1999).
[3] 中村 哲: "実音響環境に頑健な音声認識を目指して", 信学技報, EA2002-12, pp.31-36(2002).
[4] M.J.F.Gales and S.J.Young: "Robust Continuous Speech Recognition Using Parallel Model Combination", IEEE Trans. Speech and Audio Processing, Vol.4, No.5, pp.352-359, Sep.(1996).
[5] F.Martin, K.Shikano, Y.Minami and Y.Okabe: "Recognition of Noisy Speech by Composition of Hidden Markov Models", 信学技報, SP92-96, pp.9-16(1992).
[6] K.Yao, K.K.Paliwal and S.Nakamura: "Sequential Noise Compensation by A Sequential Kullback Proximal Algorithm", EuroSpeech'01, Vol.II, pp.1139-1142(2001).
[7] S.F.Boll: "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE Trans. Acoustic Speech Signal Processing, Vol.27, No.2, pp.113-120(1979).
[8] 山本寛樹, 山田雅章, 小森康弘, 大洞恭則: "推定Segmental SNRに基づく適応的Spectral Subtraction法による音声認識", 信学技報, SP94-50, pp.17-24(1994).
[9] P.Lockwood, J.Boudy and M.Blanchet: "Non-linear Spectral Subtraction (NSS) and Hidden Markov Models for Robust Speech Recognition in Car Noise Environments", ICASSP92, I-265-268(1992).
[10] M.Fujimoto, J.Ogata and Y.Ariki: "Large Vocabulary Continuous Speech Recognition under Real Environments Using Adaptive Sub-Band Spectral Subtraction", ICSLP00, Vol.I, pp.305-308(2000).
[11] C.L.Leggetter and P.C.Woodland: "Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density Hidden Markov Models", Computer Speech and Language, Vol.9, pp.171-185(1995).
[12] ELRA Web site
http://www.icp.inpg.fr/ELRA/home.html
[13] LDC Web site
http://www.ldc.upenn.edu/index.html
[14] H.G.Hirsch and D.Pearce: "The AURORA Experimental Framework for the Performance Evaluations of Speech Recognition Systems under Noisy Conditions", ISCA ITRW ASR2000, pp.18-20(2000).
[15] HTK Web site
http://htk.eng.cam.ac.uk/
[16] Description of HTK complex back-end for Aurora-2.
http://icslp2002.colorado.edu/special_sessions/aurora/
[17] HTK configuration for complex back-end for Aurora-2.
http://icslp2002.colorado.edu/special_sessions/aurora/