

## 雑音除去音声に対する特徴量抽出と MLLR適応の統合による雑音に頑健な音声認識

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、我々がこれまでに提案した雑音に頑健な音声認識手法（カルマンフィルタによる音声信号推定法と繰り返し教師無し MLLR 適応の併用）に加えて、頑健な特徴量を導入することについて検討を行った。雑音に頑健な特徴量として、Root Cepstrum 係数を用いており、音声認識に従来用いられてきた MFCC との音声認識結果の比較を行った。また、本研究では、MLLR 適応を行う際の音素クラスタ数の選択についても検討を行った。提案手法の評価は、3 種類の音楽が重畳した音声を用いた大語彙連続音声認識により行っており、提案手法により単語正解精度の改善が得られた。

キーワード：雑音に頑健な音声認識、非定常雑音、カルマンフィルタ、教師無し MLLR 適応、Root Cepstrum 係数

## Noise Robust Speech Recognition by Integration of MLLR Adaptation and Feature Extraction for Noise Reduced Speech

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 investigate a noise robust acoustic feature in our proposed noise robust speech recognition method using Kalman filtering for speech signal estimation and iterative unsupervised MLLR adaptation. For the noise robust acoustic feature, we employed root cepstral coefficients and compared the results with conventionally used MFCCs at speech recognition accuracy. Furthermore, we investigate the number of phoneme clusters in MLLR adaptation. In order to evaluate the proposed method, we carried out large vocabulary continuous speech recognition experiments under 3 types of music. As a result, the proposed method showed the significant improvement in word accuracy.

**Key words** : noise robust speech recognition, non-stationary noise, Kalman filter, unsupervised MLLR adaptation, root cepstral coefficient

## 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]。

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)[3]-[5]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)[6]-[9]の2種類が考えられる。

雑音適応の方法として、PMC(Parallel Model Combination)[3]や、NOVO(Voice mixed with NOISE)[4]に代表されるHMM合成法が提案されており、その有効性が報告されている。HMM合成法において、合成する雑音HMMの状態数、混合分布数を定常雑音の場合に比べて増加させることにより、非定常雑音への適応が可能であるといわれている。しかし、実環境下での適用を考えると、一般的には雑音HMMを学習できるのは、発話が始まるまでの区間であり、発話が始まった後の雑音の変動は学習に反映されない。このようにして学習された雑音HMMと音声HMMを合成して雑音重畳HMMを作成し、これを用いて認識を行うと、入力音声(雑音重畳音声)の時間が進むにつれ、初期の合成HMMと入力音声との間にミスマッチが生じ、認識精度に影響を与えてしまうと考えられる。この問題を解決するために、初期の合成HMMとの残差を逐次的に適応させていく手法が提案されている[5]。

HMM合成法は合成のための計算量が比較的多く、大語彙連続音声認識に使われるTriphoneモデルのHMMのように音素数、混合数の多いHMMに対して合成を行うと、非常に時間がかかってしまうという問題がある。

一方、雑音除去の観点では、Spectral Subtraction(SS)法[6]がよく知られている。しかし、SS法では雑音スペクトルの減算の際に、減算が足らずに雑音成分を残してしまったり、減算しすぎて目的とする音声のスペクトルが歪んでしまい、その結果、認識精度の低下をまねくという問題がある。また、SS法で減算する雑音スペクトルは、一般的には雑音のみが存在する区間から得られた平均スペクトルであり、非定常雑音におけるスペクトルの時間変動が考慮されていない。

以上の問題に対して、我々はこれまでに、非定常雑音

が重畳した音声における、クリーン音声(雑音の重畳していない音声)成分の時間変動の様子をモデル化し、得られたモデルに対してカルマンフィルタを適用することにより、非定常雑音が重畳した音声からクリーン音声を推定する手法について検討を行ってきた[9]。また、より高い音声認識精度を得るために、繰り返し教師無しMLLR(Maximum Likelihood Linear Regression)適応[10]を用いて、カルマンフィルタにより生じた推定誤差及び、残差雑音によるスペクトル歪みに音響モデルを適応させる方法についても検討を行ってきた[11]。本研究では、上記の音声信号推定手法、雑音適応手法に加えて、雑音に頑健な音響特徴量としてRoot Cepstrum係数[12]を用いることにより、雑音環境下における音声認識精度のさらなる改善について検討を行った。

また、MLLR適応では、音響的に類似した音素を数個のクラスタに分類した上で、適応を行っている。本研究では、MLLRにより残差雑音へ適応を行う際に、最適な音素クラスタ数についても検討を行った。

## 2 非定常雑音の除去手法

図1に本研究で用いる雑音除去手法の概念図を示す。

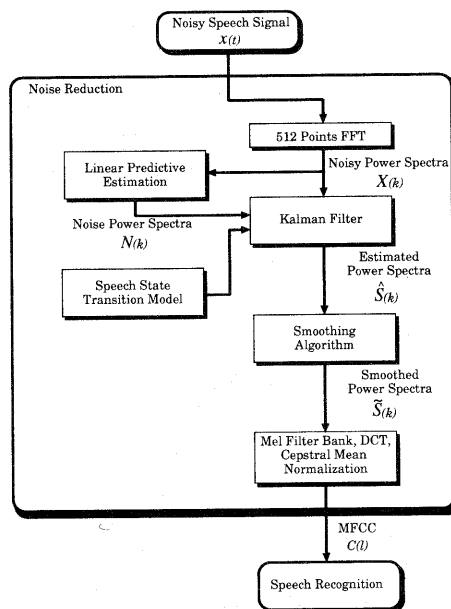


図 1: 雑音除去手法の概念図

図1において、カルマンフィルタのフィルタ方程式は、音声の時間変化モデル(雑音重畳音声に含まれる音声成分の時間変化をモデル化したもの)に基づいて定義される[9]。

ここで、 $k$  番目の短時間フレームにおいて、雑音重畳音声のパワースペクトルを  $X(k)$ 、クリーン音声のパワースペクトルを  $S(k)$ 、雑音のパワースペクトルベクトルを  $N(k)$  とすると、本研究で用いたカルマンフィルタのフィルタ方程式は、以下の式により与えられる [9].

$$\hat{S}(k) = F_{k-1}\hat{S}(k-1) + K_k \left( X(k) - F_{k-1}\hat{S}(k-1) \right) \quad (1)$$

$$K_k = Q_k (Q_k + \Sigma_{N(k)})^{-1} \quad (2)$$

$$Q_k = F_{k-1}(I - K_{k-1})Q_{k-1}F_{k-1}^T + G_{k-1}\Sigma_{W(k-1)}G_{k-1}^T \quad (3)$$

$$F_k = 1 + \Delta X^l(k) \quad (4)$$

$$\Delta X^l(k) = X^l(k+1) - X^l(k) \quad (5)$$

$$G_k = N(k) \quad (6)$$

式(1)～(6)において、 $k = 0, 1, \dots, N$  ( $N$ は最終フレーム)であり、 $\hat{S}(k)$ は $S(k)$ の推定値、 $Q_k$ は誤差の共分散行列である。また、添字 $l$ は対数パワースペクトル領域を示す。 $\hat{S}(k)$ 、 $Q_k$ の初期値はそれぞれ以下のように設定した。

$$\hat{S}(0) = 0 \quad (7)$$

$$Q_0 = 0 \quad (8)$$

式(3)の $\Sigma_{W(k)}$ は式(9)で与えられるシステム雑音 $W(k)$ の対角共分散行列であり、 $W(k)$ は平均零のガウス過程であると仮定することにより、式(11)のようにして求められる。

$$W(k) = \Delta X^l(k) - \Delta N^l(k) \quad (9)$$

$$\Delta N^l(k) = N^l(k+1) - N^l(k) \quad (10)$$

$$\Sigma_{W(k)} = W(k)W(k)^T \quad (11)$$

また、式(2)の $\Sigma_{N(k)}$ は観測雑音 $N(k)$ の対角共分散行列であり、 $W(k)$ 同様、平均零のガウス過程であると仮定して、以下のようにして求めた。

$$\Sigma_{N(k)} = N(k)N(k)^T \quad (12)$$

式(9)、(12)において、 $W(k)$ 、 $\Sigma_{N(k)}$ の計算に必要となるベクトル $N(k)$ は、12次の線形予測法により推定している。

### 3 Root Cepstrum 係数を用いた特徴抽出

音声認識の特徴量として、MFCCが一般的に用いられており、MFCC特徴ベクトルの第 $i$ 要素を $MFCC_i$ 、第 $j$ 番目のメルフィルタバンク出力(次元数 $N$ )を $s_j$ 、DCT

係数を $DCT_{ij}$ としたとき、MFCCは式(13)のように計算される。

$$MFCC_i = \sqrt{\frac{2}{N}} \sum_{j=0}^{N-1} \log(s_j) DCT_{ij} \quad (13)$$

MFCCは、雑音の存在しないクリーンな音声において、効果的な特徴量であるが、雑音が重畳した音声の場合、認識率を低下させてしまう。本研究で用いている雑音除去法では、カルマンフィルタで推定された音声信号に、推定誤差による残差雑音が多く含まれていることが予備実験によりわかっている [11]。このことから、雑音除去法により推定された音声信号に対して、MFCCを用いて特徴抽出すると、残差雑音の影響を大きく受けしまい、音声認識率を低下させるという問題が生じる。この問題に対して、本研究では、比較的雑音に頑健なパラメータとされるRoot Cepstrum 係数(RTCC)[12]の特徴量として利用することを試みた。特徴ベクトルの第 $i$ 要素を $RTCC_i$ としたとき、RTCCは式(14)のように計算される。

$$RTCC_i = \sqrt{\frac{2}{N}} \sum_{j=0}^{N-1} (s_j)^\gamma DCT_{ij} \quad (14)$$

式(14)において、 $\gamma$ は一般に $0 < \gamma \leq 1$ の実数を用いられしており、本研究では $\gamma = 0.08$ を用いてRTCCを算出している。

## 4 実験

2, 3で述べた手法を用いて、雑音下における大語彙連続音声認識実験を行った。

### 4.1 実験条件

評価用データには、IPA-98-TestSetのうち、男性23名が発声したデータ100文を用いている。また、重畳させる非定常雑音は音楽であり、ピアノソロ曲3曲(Piano1, Piano2, Piano3)を用いた。雑音の重畳はそれぞれの音楽データからランダムに切り出した100区間分のデータを、式(15)を用いて $SNR(= 20, 10, 0\text{dB})$ を調整した後、それぞれ音声データに計算機を用いて重畳させた。

$$x(t) = s(t) + \frac{Pow_s}{10^{SNR/20} Pow_n} \cdot n(t) \quad (15)$$

ここで、 $x(t)$ 、 $s(t)$ 、 $n(t)$ はそれぞれ雑音重畳音声、音声、雑音を表し、 $Pow_s$ 、 $Pow_n$ はそれぞれ音声、雑音のRMS(Root Mean Square)パワーを表す。

音響モデルには、話者独立なmonophone HMMを用いた。HMMの学習には、日本音響学会新聞記事読み上げ音声コーパスのうち、男性話者137人分の21782発

話を用いており、それぞれのデータに対して Cepstrum Mean Normalization(CMN)を行っている。音響分析の条件、HMMの構造を表1, 2に示す。

表 1: 音響分析条件

|                     |                                                             |
|---------------------|-------------------------------------------------------------|
| 標準化周波数              | 16kHz                                                       |
| 高域強調                | $1 - 0.97z^{-1}$                                            |
| 特徴パラメータ<br>(学習, 認識) | 12次MFCC及び12次RTCC<br>+ log Power + $\Delta$ + $\Delta\Delta$ |
| 特徴パラメータ<br>(雑音除去)   | 512点FFTスペクトル                                                |
| 分析区間長               | 20ms                                                        |
| 分析周期                | 10ms                                                        |
| 時間窓                 | Hamming Window                                              |
| SNR                 | 20,10,0dB                                                   |

表 2: 音素HMMの構造

|     |                   |
|-----|-------------------|
| 状態数 | 5状態3ループ           |
| 混合数 | 12                |
| 音素数 | 41                |
| タイプ | Left-to-Right HMM |

言語モデルには、1st-passにbigram, 2nd-passにtrigramを用いており、IPAモデル98年度版のうち、語彙数20k, cut-offはbigram, trigramそれぞれに対して4-4のモデルを用いている。言語モデルの学習データは、毎日新聞記事75ヶ月分である。

## 4.2 実験結果

表3～5にそれぞれの手法による認識結果を示す。それぞれの手法において、上段は特徴量にMFCCを用いた場合の単語正解精度、下段はRTCCを用いた場合の単語正解精度を示す。ここで、クリーンな音声での単語正解精度は、MFCCでは86.49%であり、RTCCでは85.29%である。

表 3: 認識結果 (Piano 1)(%)

| SNR    |      | 20dB  | 10dB  | 0dB   |
|--------|------|-------|-------|-------|
| 雑音除去無し | MFCC | 79.52 | 61.57 | 20.04 |
|        | RTCC | 80.50 | 66.48 | 25.97 |
| 雑音除去有り | MFCC | 80.06 | 67.79 | 34.69 |
|        | RTCC | 80.72 | 70.96 | 38.68 |

表 4: 認識結果 (Piano 2)(%)

| SNR    |      | 20dB  | 10dB  | 0dB   |
|--------|------|-------|-------|-------|
| 雑音除去無し | MFCC | 79.62 | 56.44 | 20.04 |
|        | RTCC | 78.98 | 60.87 | 27.71 |
| 雑音除去有り | MFCC | 78.95 | 61.57 | 31.64 |
|        | RTCC | 78.50 | 65.06 | 38.24 |

表 5: 認識結果 (Piano 3)(%)

| SNR    |      | 20dB  | 10dB  | 0dB   |
|--------|------|-------|-------|-------|
| 雑音除去無し | MFCC | 79.45 | 61.95 | 21.81 |
|        | RTCC | 80.42 | 65.85 | 28.32 |
| 雑音除去有り | MFCC | 80.27 | 66.01 | 33.48 |
|        | RTCC | 80.52 | 69.75 | 38.55 |

それぞれの表において、特徴量にRTCCを用いることにより、MFCCを用いた場合に比べて、単語正解精度の改善が得られたが、全体的に雑音除去による改善量は小さい。それぞれの結果を解析した結果、雑音除去により、置換誤り及び、脱落誤りは減少したが、挿入誤りが増加したことがわかっている [9]。

挿入誤りが増加した理由として、図1に示した手法では、ベクトル  $N(k)$  の推定を線形予測法を用いて行っているが、時間  $k$  が経過するにつれて、推定誤差が大きくなり、十分な推定精度が得られなかったためであると考えられる。この雑音成分の推定精度の低さが音声成分の推定精度に影響を与え、認識時に音響スコアが、低くなってしまったと考えられる。また、デコーディングがこの音響スコアに基づいて行われるため、誤った単語が連結されてしまい、挿入誤りが増加したと考えられる。

## 5 雑音除去とモデル適応の併用

### 5.1 繰り返し教師無し MLLR 適応

4の実験において、ベクトル  $N(k)$  の推定誤差が大きかったために、カルマンフィルタによる推定信号  $\hat{S}(k)$  に推定誤差及び、残差雑音によるスペクトル歪みが発生してしまっている。このスペクトル歪みの影響により、2の雑音除去法では、単語正解精度の十分な改善が得られなかった。この問題を解決するために、繰り返し教師無し MLLR(Maximum Likelihood Linear Regression) 適応 [10] を用いて、音響モデルを推定誤差及び、残差雑音によるスペクトル歪みに適応させた。ここで、教師無し MLLR 適応を適用するためには、適応に用いる音声データの音素ラベルが必要となる。本研究では、カルマンフィルタにより推定された音声信号を用いて音素認識を行い、音素認識の結果を音素ラベルとして用いている。

### 5.2 MLLR 適応における音素クラスタ数

MLLR 適応では、音響的に類似した音素を数個のクラスタに分類し、各音素クラスタ毎に適応が行われる。また、音素認識等により得られた音素ラベルに基づいて、適応データに対して時間方向に音素列の Viterbi アライメント等をとることにより、適応データから各音素区間を取り出し、各クラスタ毎に適応を行う。本研究では、カ

ルマンフィルタにより推定された音声信号に対して音素認識を行い、その結果を音素ラベルとして使用している。

しかし、前述のように推定信号には推定誤差が含まれており、数字の上での音素認識率が比較的良好であっても、Viterbiアライメントをとったときに、各音素境界が大きすぎてしまうと考えられる。このため、MLLRによりあるクラスタを適応する際に、実際にはクラスタ内には含まれない音素の区間であるにも関わらず、音素境界のずれにより誤ってその区間を適応データとして使用してしまい、適応の精度を低下させることが考えられる。

この問題を解決するために、本研究では、繰り返し適応において、1回目の適応を行う際の音素クラスタ数を1とした。つまり、音素のクラスタ分類を行わないということであり、音素境界の誤りが生じて、無関係なクラスタに誤って適応データを与えてしまうという問題を回避できると考えられる。

また、1回目の適応により得られた適応HMMを用いて、Viterbiアライメントをとることにより、1回目の適応時に比べて音素境界のずれを補正できると考えられる。音素のクラスタ分類は、2回目以降の繰り返し適応の際に行っており、補正された音素境界情報を用いることにより、1回目の適応時に比べて各クラスタに正しい区間境界を持った適応データを与えることができる。

## 6 繰り返し教師無し MLLR 適応を用いた実験結果

5.2で述べた、音素クラスタ数選択の効果を示すために、最大3回の繰り返し教師無しMLLR適応を、表6に示すような2つのパターンで行った。適応パターンAは、1回目にクラスタ数1で適応を行い、その後の適応で、クラス多数を増加させる。一方、適応パターンBは常にクラスタ数16で適応を行う。

表 6: 繰り返し適応時の音素クラスタ数変化パターン

| 適応回数 | 適応パターンAのクラスタ数 | 適応パターンBのクラスタ数 |
|------|---------------|---------------|
| 1    | 1             | 16            |
| 2    | 8             | 16            |
| 3    | 16            | 16            |

表7～9に、繰り返し教師無しMLLR適応を用いた認識結果を示す。本研究では、適応用データには評価用データ100全文全てを用いている。それぞれの表において、'雑音除去無し'は雑音が重畳した音声を用いて、音素ラベル作成及び、HMMの適応を行い、雑音が重畳した音声を認識した結果を示している。一方、'雑音除去有り'は2、3で述べた音声信号推定法により推定した音声を用い

て、音素ラベル作成及び、HMMの適応を行い、推定した音声を認識した結果を示している。

それぞれの表において、2回の繰り返しで、認識精度の改善がほぼ飽和している。また、それぞれの雑音環境において、適応を行わなかった場合(適応回数0)に比べて、認識精度の改善が得られている。特に、雑音の影響が強い0dBの環境において、大幅な認識精度の改善が得られ、全体的にMFCCよりもRTCCの方が高い精度を示している。

ここで、適応パターンAと適応パターンBの結果を比較すると、全体的に適応パターンAの結果の方が高い結果を示しており、この傾向は低SNRになるほど顕著に現れている。5.2でも述べたように、特に低SNRの環境では、適応時のViterbiアライメント等による、各音素境界の時間情報の信頼性が低く、誤った音素区間を音素クラスタに適応データとして与えてしまう。しかし、適応パターンAのように、クラスタ数1で最初の適応を行った場合、音素境界の誤りが生じて、全ての適応データは1つのクラスタにのみ与えられるので、無関係なクラスタに誤った適応データを与えてしまうという問題を回避することができる。このことから、音素境界の時間情報の信頼性が低い場合は、音素クラスタ数を1のような小さい値に設定することにより、精度の高い適応を行うことができると言える。

## 7 おわりに

本研究では、雑音除去と雑音適応に加えて、雑音に頑健な特徴量を導入することにより、認識精度の改善が得られることを示した。また、繰り返し教師無しMLLR適応を行う際の、音素クラスタ数についても検討し、特に低SNRでは、音素クラスタ数を1のような小さい値に設定することにより、適応及び、認識精度の改善が得られる事を示した。

今後、雑音除去精度の改善、雑音に一層頑健な特徴量について検討する予定である。また、今回の実験では、繰り返し教師無しMLLR適応の適応データとして、100文章の音声データを用いていたが、今後、少ない適応データで、高速に適応を行う手法について検討を行う予定である。

## 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] 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).

表 7: 繰り返し教師無し MLLR 適応を用いた認識結果 (Piano 1)(MFCC (%) / RTCC (%))

|        | 適応回数 | 適応パターン A             |                      |                      | 適応パターン B             |                      |                      |
|--------|------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
|        |      | 20dB                 | 10dB                 | 0dB                  | 20dB                 | 10dB                 | 0dB                  |
| 雑音除去無し | 0    | 79.52 / 80.50        | 61.57 / 66.48        | 20.04 / 25.97        | 79.52 / 80.50        | 61.57 / 66.48        | 20.04 / 25.97        |
|        | 1    | 83.51 / <b>84.78</b> | 75.90 / 77.62        | 44.13 / <b>52.69</b> | 83.51 / 83.39        | 74.06 / 76.66        | 41.92 / <b>50.92</b> |
|        | 2    | 82.70 / 84.59        | <b>78.50</b> / 77.11 | 44.58 / 52.19        | <b>83.70</b> / 81.74 | <b>78.12</b> / 74.13 | 40.76 / 49.71        |
|        | 3    | 81.98 / 82.63        | 77.11 / 76.73        | 48.07 / 51.24        | 83.39 / 81.42        | 74.32 / 72.86        | 41.39 / 48.95        |
| 雑音除去有り | 0    | 80.06 / 80.72        | 67.79 / 70.96        | 34.69 / 38.68        | 80.06 / 80.72        | 67.79 / 70.96        | 34.69 / 38.68        |
|        | 1    | 82.69 / 83.26        | 76.66 / 79.07        | 57.51 / 61.57        | 81.99 / <b>83.20</b> | 76.35 / 78.36        | 57.07 / 57.58        |
|        | 2    | 82.88 / <b>84.15</b> | 77.87 / <b>79.63</b> | 62.51 / 64.05        | 81.99 / 82.88        | 76.98 / <b>78.74</b> | 59.73 / <b>61.00</b> |
|        | 3    | 82.63 / 83.75        | 77.17 / 77.43        | 62.71 / <b>65.00</b> | 81.10 / 80.72        | 77.41 / 77.49        | 57.64 / 58.21        |

表 8: 繰り返し教師無し MLLR 適応を用いた認識結果 (Piano 2)(MFCC (%) / RTCC (%))

|        | 適応回数 | 適応パターン A             |                      |                      | 適応パターン B             |                      |                      |
|--------|------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
|        |      | 20dB                 | 10dB                 | 0dB                  | 20dB                 | 10dB                 | 0dB                  |
| 雑音除去無し | 0    | 79.62 / 78.98        | 56.44 / 60.87        | 20.04 / 27.71        | 79.62 / 78.98        | 56.44 / 60.87        | 20.04 / 27.71        |
|        | 1    | <b>83.01</b> / 82.88 | 67.98 / 74.13        | 39.06 / 42.49        | <b>83.13</b> / 81.80 | 65.44 / 71.66        | 37.41 / 42.37        |
|        | 2    | 81.36 / 82.05        | 71.15 / <b>75.69</b> | 40.90 / <b>42.87</b> | 82.56 / 79.96        | 71.53 / 70.96        | 37.67 / 41.92        |
|        | 3    | 79.96 / 81.29        | 69.37 / 74.25        | 42.17 / 42.04        | 82.37 / 79.49        | <b>71.91</b> / 70.67 | 40.84 / <b>42.55</b> |
| 雑音除去有り | 0    | 78.95 / 78.50        | 61.57 / 65.06        | 31.64 / 38.24        | 78.95 / 78.50        | 61.57 / 65.06        | 31.64 / 38.24        |
|        | 1    | 81.88 / <b>83.45</b> | 73.87 / 73.24        | 52.52 / 52.76        | 80.91 / <b>82.94</b> | 70.20 / 73.18        | 46.99 / 50.10        |
|        | 2    | 81.61 / 81.99        | 73.87 / <b>75.46</b> | 53.46 / <b>57.20</b> | 81.23 / 82.05        | 73.49 / <b>74.49</b> | 52.00 / 52.25        |
|        | 3    | 80.28 / 82.05        | 73.43 / 74.25        | 53.14 / 56.10        | 78.25 / 80.66        | 72.23 / 72.86        | <b>52.44</b> / 50.98 |

表 9: 繰り返し教師無し MLLR 適応を用いた認識結果 (Piano 3)(MFCC (%) / RTCC (%))

|        | 適応回数 | 適応パターン A             |                      |                      | 適応パターン B             |                      |                      |
|--------|------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
|        |      | 20dB                 | 10dB                 | 0dB                  | 20dB                 | 10dB                 | 0dB                  |
| 雑音除去無し | 0    | 79.45 / 80.42        | 61.95 / 65.85        | 21.81 / 28.32        | 79.45 / 80.42        | 61.95 / 65.85        | 21.81 / 28.32        |
|        | 1    | 83.39 / <b>83.83</b> | 75.75 / <b>78.19</b> | 47.50 / 53.01        | 83.45 / 82.56        | 74.06 / 75.59        | 47.05 / 49.90        |
|        | 2    | 82.25 / 83.07        | 77.36 / 77.62        | 50.22 / 51.68        | <b>83.70</b> / 81.04 | <b>77.74</b> / 74.32 | 42.85 / <b>50.60</b> |
|        | 3    | 80.28 / 82.12        | 76.35 / 77.05        | <b>53.08</b> / 49.97 | <b>83.70</b> / 82.98 | 77.17 / 74.25        | 47.75 / 49.27        |
| 雑音除去有り | 0    | 80.27 / 80.52        | 66.01 / 69.75        | 33.48 / 38.55        | 80.27 / 80.52        | 66.01 / 69.75        | 33.48 / 38.55        |
|        | 1    | 81.80 / <b>83.70</b> | 76.26 / 76.73        | 61.76 / 61.51        | 82.50 / <b>83.20</b> | 74.57 / 77.17        | 58.47 / 59.48        |
|        | 2    | 82.24 / 81.48        | 76.54 / <b>77.49</b> | 60.88 / <b>62.97</b> | 81.86 / 82.82        | 74.30 / <b>77.36</b> | 60.05 / <b>62.84</b> |
|        | 3    | 81.80 / 82.18        | 76.60 / 76.92        | 60.94 / 62.08        | 80.98 / 82.88        | 74.25 / 76.35        | 58.02 / 58.91        |

- [4] 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).
- [5] K.Yao, K.K.Paliwal and S.Nakamura: "Sequential Noise Compensation by A Sequential Kullback Proximal Algorithm", *Proc. of EuroSpeech'01*, Vol.II, pp.1139-1142(2001).
- [6] 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).
- [7] D.C.Popescu, I.Zejković: "Kalman Filtering of Colored Noise for Speech Enhancement", *Proc. of ICASSP'98*, Vol.II, pp.997-1000(1998).
- [8] Z.Goh, K.Tan and B.T.G.Tan: "Kalman-Filtering Speech Enhancement Method Based on Voiced-Unvoiced Speech Model", *IEEE Trans. Speech and Audio Processing*, Vol.7, No.5, pp.510-524, Sep.(1999).
- [9] M.Fujimoto and Y.Ariki: "Continuous Speech Recognition under Non-stationary Musical Environments Based on Speech State Transition Model", *Proc of ICASSP'01*, Vol.I, pp.297-300(2001).
- [10] 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).
- [11] M.Fujimoto and Y.Ariki: "Speech Recognition under Musical Environments Using Kalman Filter and Iterative MLLR Adaptation", *Proc of EuroSpeech'01*, Vol.III, pp.1879-1882(2001).
- [12] Umit Yapanel, Jhon H.L. Hansen, Ruhi Sarikaya and Bryan Pellom: "Robust Digit Recognition in Noise: An Evaluation Using AURORA Corpus", *Proc. of EuroSpeech'01*, Vol.I, pp.209-212(2001).