音声言語情報処理 44－4
(2002. 12. 19)

雑音に頑健な音韻モデルと教師なし話者適応

山出 慎吾† 李 晃伸† 猿渡 洋† 鹿野 清宏†

† 奈良先端科学技術大学院大学 情報科学研究科
〒 630-0912 奈良県生駒市高山町 8916-5
E-mail: †{shing-y, ri, sawatari, shikano}@is.aist-nara.ac.jp

あらまし 実環境において頑健に音声認識を行うためには，音韻モデルを環境や話者に対して適応させることが重要である．特に実用化を考慮した場合，環境雑音の変動や話者の交代に迅速に対応できることが必要となる．本稿では，まず雑音下の入力音声に対しスペクトルサブトラクションを施した後，任意の小量の雑音を重畳することにより，雑音雑音の影響を低減するアルゴリズムを提案する．さらに，提案手法を十分統計量に基づく教師なし話者適応アルゴリズムに適用する．従来は対象とする環境ごとに音声データベースに雑音を重畳して十分統計量を計算する必要があったが，提案手法では各雑音の種類やSNRの変化をスペクトルサブトラクションおよび雑音の重畳により打ち消すため十分統計量を再計算する必要がなく，どのような環境においても任意の一発声文で，高速に音韻モデルの教師なし話者適応が行える．提案法をオフィス，車内，展示会場，人混みの4種類の環境において，2万語のディクテーションタスクで認識実験を行ったところ，提案手法により適応した音韻モデルの平均認識率は，雑音環境ごとにマッチさせた従来の環境・話者適応モデルと比較してほぼ同程度の認識性能を示し，さらに雑音の変動に対する頑健性が示された．さらに教師あり適応であるMLLR法との比較も報告する．

キーワード 耐雑音音声認識，話者適応，スペクトルサブトラクション，十分統計量

Noise Robust Speech Recognition Applied to Unsupervised Speaker Adaptation

Shingo YAMADE†, Akinobu LEE†, Hiroshi SARUWATARI†, and Kiyohiro SHIKANO†

† Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama-cho, Ikoma, Nara, 630-0912, Japan
E-mail: †{shing-y, ri, sawatari, shikano}@is.aist-nara.ac.jp

Abstract Noise and speaker adaptation techniques are essential to realize robust speech recognition in real noisy environments. We proposed that a noise robust speech recognition is implemented by superimposing a small quantity of noise data on spectral subtracted input speech. We also apply this noise robust speech recognition to the unsupervised speaker adaptation algorithm based on HMM sufficient statistics in different noise environments. According to spectral subtraction and nois superimposition, our proposed algorithm can make robust against the change of noises and SNR, and adapt quickly without calculating HMM sufficient statistics from noise matched acoustic models. We evaluate successfully our proposed algorithm with 20 k dictation task using four kinds of noises. The recognition experiments show that our proposed method increases the robustness against different noises significantly . We also compared our proposed method with unsupervised MLLR adaptation.

Key words Noise Robust Speech Recognition, Speaker Adaptation, Spectral Subtraction, HMM Sufficient Statistics

1. はじめに

実環境において頑健に音声認識を行うためには，音韻モデルの環境・話者適応は必要である．しかしながら実環境では，様々な雑音が存在するため，その全ての環境において適応効果の高い環境・話者適応を行うことは困難である．我々は，現在までに環境および話者適応法について研究を進めてきている [1]〜[3]．これまでの環境・話者適応手法 [3] では適応に用いる十分統計量を用意するためにあらかじめ雑音マッチドモデルを作成する必要があった．しかし，雑音マッチドモデルは作成に膨大な計算量を必要とする．また，スペクトルサブトラクション (以下 SS) [3], [4] を施すと SNR は大きく改善するが，一般にフロアリングなどの処理を行っているため，SS 後の音声スペクトルは，元の雑音スペクトルの特性を残している．ゆえに十分に SS の性能を発揮するには，その特性を考慮する必要がある．

そこで本稿では，音韻モデルの異なる雑音下における頑健性を向上させるために，SS 後の入力音声に対し小量の雑音を重畳し，SS 後の残差雑音等の歪みの影響を低減させる手法を提案する．重畳する雑音はシステム側にとっては既知の雑音であるため，あらかじめ保持しておいた雑音マッチドモデルで認識することで高い認識性能が期待される．またこの手法と我々の既提手法である教師なし話者適応アルゴリズム [2] とを統合することで，さらに高い効果が得られることが期待される．従来法 [3] は各雑音環境ごとに雑音マッチドモデルを作成する必要があり，雑音が変動した場合に認識精度が低下してしまう可能性があるため，雑音マッチドモデルを作成し直す必要があった．しかし統合した適応アルゴリズムにより，雑音環境が変動しても雑音マッチドモデルを改めて作成する必要がなく，高速に環境および話者に適応することが可能となる．

以下，2 章では提案する雑音に頑健なアルゴリズムについて述べ，オフィス，展示会場，車内，人混み [5] の雑音環境に対する 2 万語のディクテーションタスク [6], [7] において評価を行う．3 章では提案手法と教師なし話者適応アルゴリズムを統合した適応アルゴリズムについて述べ，評価実験結果について述べる．さらに，環境・話者適応手法として一般的に広く使用されている教師あり MLLR (Maximum Likelihood Linear Regression) [8] と提案手法とを比較する．4 章ではまとめを述べる．

2. 既知雑音の重畳を用いた雑音に頑健な音声認識アルゴリズム

実環境における連続音声認識においては，雑音に対する頑健性が必要である．広く用いられている雑音への対処法としては，目的雑音を音声データベースに重畳してモデルを構築する雑音マッチドモデルが高い精度を示す．しかし実環境においては様々な雑音環境が存在し，あらかじめあらゆる種類の雑音データを保持しておくことは不可能である．そこで我々は，SS と雑音重畳を用いた異なる雑音環境に頑健な音声認識アルゴリズムを提案する．

2.1 スペクトルサブトラクション

SS [4] は，雑音が定常であることを仮定して，非音声区間の信号より雑音の特徴量を推定しておき，雑音混じりの音声の特徴量から雑音を取り除き，元の音声信号を推定する信号処理である．時刻 t の真の音声信号を $s(t)$，雑音信号を $n(t)$ とすると，観測される雑音混じりの信号 $y(t)$ は，次のように表される．

$$y(t) = s(t) + n(t) \tag{1}$$

ここで，窓の位置を m で表した短時間分析による両辺のフーリエ変換をとると次の式が得られる．

$$Y(f,m) = S(f,m) + N(f,m) \tag{2}$$

ただし，$Y(f,m)$，$S(f,m)$，$N(f,m)$ は，周波数 f の複素スペクトルを表す．振幅スペクトルは $Y(f,m)$ から推定雑音を減算したものとし，位相は入力信号のものを使用する．この方法により推定される雑音信号は次のように表される．

$$\hat{S}(f,m) = \left||Y(f,m)|^2 - \alpha E_m\left[|N(f,m)|^2\right]\right|^{1/2} \cdot e^{j\arg\left(Y(f,m)\right)} \tag{3}$$

ただし，$E_m\left[|N(f,m)|^2\right]$ は非音声区間で推定した雑音信号，α は減算調節パラメータである．また，推定された音声の振幅スペクトルが負である場合やフレーム単位の処理により発生するミュージカルノイズを回避するため，本研究では次に示す式のように任意の係数 $A(0 < A < 1)$ を用いてフロアリング処理を行う．

$$\hat{S}(f,m) = Y(f,m) \cdot A$$

$$\left(\text{ただし } |Y(f,m)|^2 - \alpha E_m\left[|N(f,m)|^2\right] < 0\right) \tag{4}$$

2.2 既知雑音の重畳

一般に SS による改善 SNR は 10 dB 程度と言われているが，残差雑音により SS 後のスペクトルに元の雑音スペクトルの特性を残している．またフロアリング処理による歪みも発生する．そこで我々は，SS 後の入力音声をより精度よく認識するために，目的雑音を重畳した音声データベースに SS を施して作成した雑音マッチドモデル (SS Matched) を用いた [3]．この SS Matched モデルは，SS により，各環境間の差異を減少させることができるため，環境が異なってもある程度頑健にすることが可能である．しかし，SS による残差雑音が大きく残る環境においては，かえって認識率の低下を招いてしまう場合があった．また，foward-backward アルゴリズムにより再学習を行うため，計算時間が膨大である．そこで本稿では，この SS Matched を使用するかわりに，SS 後の入力音声に小量の任意の雑音を重畳し，その音声をその重畳した雑音のマッチドモデルを用いて認識を行うアルゴリズムを提案する．SS 後の残差雑音の影響を，雑音を重畳することで低減することが可能である．また，あらかじめ準備する雑音マッチドモデルは一種類でよいため，新たに雑音マッチドモデルを作成し直す必要が無く，認識精度を高く保つことが可能となる．

表 1　オフィス雑音を重畳した場合の認識率
Table 1　Word accuracy for office noise superimposition

Input	clean	30 dB SNR			25 dB SNR		
PTM	clean	30 dB	25 dB	20 dB	30 dB	25 dB	20 dB
(SI)	91.1	89.9	89.3	86.8	86.2	86.3	86.0

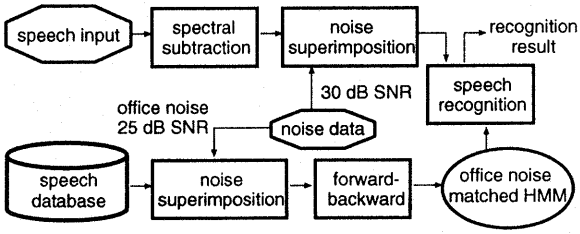

図 1　SS と雑音重畳に基づく雑音に頑健な認識アルゴリズム
Fig. 1　Noise robust speech recognition algorithm by spectral subtraction and noise superimposition

まず，予備実験として，タスクは 2 万語のディクテーションタスク [6], [7] において，重畳する雑音について検討を行った．重畳する雑音としては比較的定常で [3] において適応の効果の高かった主に計算機ノイズからなるオフィス雑音を使用する．clean 環境の場合と，オフィス環境雑音の 30 dB，25 dB，20 dB SNR マッチドモデルを用いた場合の認識実験の結果を表 1 に示す．表中の認識率は，PTM (Phonetic Tied Mixture) [9] モデルによる単語正解精度である．

表 1 より，30 dB SNR の入力に対し 30 dB の雑音マッチドモデルで認識した場合の認識率は，clean 条件の場合とほぼ差がなかった．また，同じ 30 dB SNR の入力に対して 25 dB の雑音マッチドモデルで認識した場合も，30 dB の場合とほぼ同程度の認識精度が得られた．この結果より，SS 後の入力波形に 30 dB のオフィス雑音を重畳することとする．

提案する SS と雑音重畳に基づく雑音に頑健な認識アルゴリズムを図 1 に示す．まずオフラインで学習用音声データベースに対しオフィス雑音を重畳する．ただしこのとき，入力音声の SS 後に残差雑音が生じるため，音韻モデルの SNR より悪化する事が考えられる．この点を考慮して，学習用データベースには SNR が 25 dB の雑音を重畳する．その後 forward-backward アルゴリズムを用いて雑音マッチドモデルを作成しておく．認識する際，まず雑音下の入力音声に対し SS を施す．その後 30 dB SNR のオフィス雑音を重畳し，その音声を認識器側に送り，あらかじめ作成しておいた雑音マッチドモデルで認識を行う．手法の処理手順をまとめると以下の通りである．

1. SNR 25 dB のオフィス雑音を音声データベースに重畳し，雑音マッチドモデルを作成する．
2. 入力評価音声に対し SS を施す．
3. SNR 30 dB のオフィス雑音を重畳し，1. のモデルで認識を行う．

以上により，環境が変動しても音韻モデルを再学習することなく頑健に認識することができ，またオフィス以外の雑音に対しても頑健である．

表 2　雑音の種類
Table 2　Types of noise

office	オフィス環境における居室雑音 (マイク収録)
car	高速道路走行中の車内雑音 (電子協騒音 DB No.1)
booth	展示会場内ブース雑音 (電子協騒音 DB No.3)
crowd	人混みにおける雑音 (電子協騒音 DB No.10)

表 3　実験条件
Table 3　Experimental condition

データベース	JNAS [6](306 人，150 文章/一人)
標本化/量子化	16 kHz/16 bit
窓シフト長	10 msec
特徴量	MFCC(12 次元)，ΔMFCC，Δ パワー
言語モデル	20 k(新聞記事)
音韻モデル	PTM [9](2000 状態，64 混合)
学習用データ	260 人 (150 文章/一人)
評価用データ	46 人 (200 文章)
デコーダ	Julius ver.3.1

2.3　提案法の評価

提案手法の有効性を検証するために，種々の雑音環境 2 について，提案した SS と雑音重畳に基づく認識アルゴリズムの評価実験を行った．雑音環境の種類としては，2.2 節で用いたオフィス雑音，および電子協騒音データベース [5] から車内，展示会場，人混みの雑音の計 4 種を用いる．提案法を，雑音マッチドモデルおよび SS Matched モデルと比較する．

2.3.1　実験条件

実験条件を表 3 に示す．学習用音声データベースは JNAS [6] の音声データベースを用いる．JNAS データベースは，男性 153 人，女性話者 153 人の計 306 人で構成されており，各話者ごとに 50 文の音素バランス文と約 100 文の新聞記事読み上げ文を持つ．今回の実験では 306 人中男性 130 人，女性 130 人の計 260 人を学習用話者とし，残る男性 23 人，女性 23 人の計 46 人を評価用話者として用いた．また学習用データ数は，各話者につき音素バランス文 50 文，新聞読み上げ 100 文である．評価用データしては各話者につき 4，5 文章，計 200 文章である．サンプリング条件は 16 kHz，16 bit，特徴量は窓シフト長 10 ms で分析した 12 次元の MFCC (Mel-frequency cepstral coefficient) と ΔMFCC，Δ パワーを用いる．言語モデルは新聞記事から構築した語彙数 20 k の 3-gram を使用し，デコーダは Julius を用いる．適応に使用する初期モデルとして，260 人の学習用話者から学習した 2000 状態 64 混合の PTM [9] 不特定話者モデルを使用する．

SS に用いる減算調節パラメータ α は 2.0，フロアリング係数 A は 0.5 とする．SS 後の重畳雑音は 30 dB SNR のオフィス雑音を用いる．

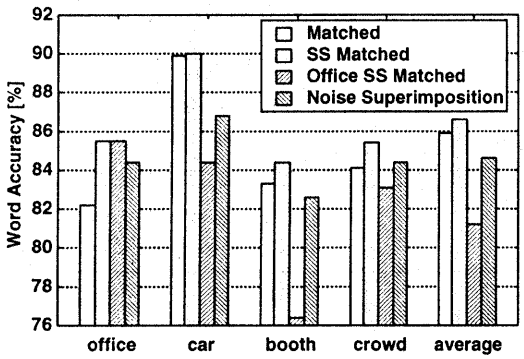

図 2 雑音に頑健な認識アルゴリズムの評価結果 (20 dB 入力)

Fig. 2 Evaluation result for noise robust speech recognition algorithm (input : 20 dB SNR)

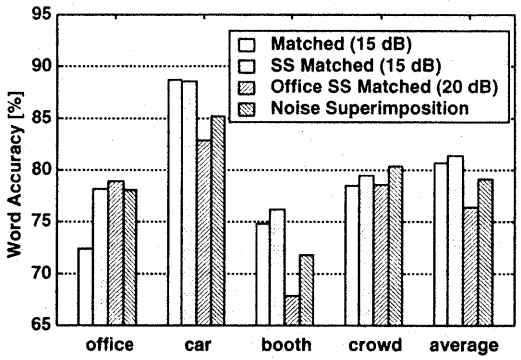

図 3 雑音に頑健な認識アルゴリズムの評価結果 (15 dB 入力)

Fig. 3 Evaluation result for noise robust speech recognition algorithm (input : 15 dB SNR)

2.3.2 実験結果

入力音声の SNR を 20 dB としたときの，提案法の評価結果を図 2 に示す．グラフにおいて縦軸は単語正解精度 (%)，横軸は各環境雑音および実験条件を示す．

図 2 中の Matched は，目的雑音を重畳した雑音重畳データベースから再学習により作成した雑音マッチドモデルでの認識率を示す．SS Matched は，目的雑音を重畳した後，SS を施したデータで学習したモデルでの認識率，Office SS Matched は，あらかじめ作成したオフィス環境雑音の SS Matched モデルでの認識率を示している．Noise Superimposition は提案法による認識率である．図 2 から，car，booth，crowd において，Office SS Matched は，入力音声の SS 後に生じる残差雑音の特徴がモデル側と異なるために，SS Matched よりも認識率が低下する．しかし，提案法の Noise Superimposition は，Office SS Matched よりも認識率が向上した．これは小量の雑音を重畳することで残差雑音の影響を低減することができるためである．なお，office では SS Matched が入力と学習時音声が同一の雑音であるのに対し，Noise Superimposition は SS 後の音声にさらに雑音を重畳しているが，1%程度の認識率の低下で収まっている．

car，booth，crowd における単語正解精度の平均値は，SS Matched の 81.6%から Noise Superimposition の 84.6%まで向上した．以上により，提案手法である Noise Superimposition は雑音が異なっても同一モデルで高い認識精度が得られることが示された．また，PMC (Parallel Model Combination) について以前調査したもの [10] と比較しても，より良い結果となった．

さらに SNR が異なる環境での性能を調べるために，より SNR の低い 15 dB SNR の入力に対して，頑健性を評価した．実験結果を図 3 に示す．結果は同様の傾向となり，car，booth，crowd における単語正解精度の平均値は，SS Matched モデルの 81.4%に対し，20 dB SNR の Office SS Matched モデルで 76.4%であったものが，提案手法により 79.1%と大きく改善された．以上の結果より，雑音の種類の変化や SNR の変動に対しても，雑音マッチドモデルや環境適応等の処理を行うことなく，提案手法は高い認識精度を保つことが可能である．

3. 種々の雑音環境における教師なし話者適応

2 章で提案した SS と雑音重畳に基づく認識アルゴリズムを，十分統計量に基づく教師なし話者適応法 [2] と統合する．手法の概要を図 4 に示す．従来の十分統計量を用いた適応アルゴリズム [3] では，十分統計量を計算する際，適応対象の環境ごとに雑音マッチドモデル (SS Matched) が必要であったが，本提案手法と統合することで，あらかじめ 25 dB のオフィス雑音マッチドモデルから十分統計量を計算しておくことができる．また統合後の提案手法は，任意の一文章のみを用いて高速に話者に適応することが可能であるため，入力話者にかかる負担が少なく，話者の交替にも迅速に対応することができる．このように，両手法を組み合わせることで，環境の変動および話者の交替の双方に頑健な音声認識が実現できる．

3.1 適応アルゴリズム

提案手法を統合した適応アルゴリズムは，以下の 6 段階の手順からなる．

1. あらかじめ，25 dB SNR でオフィス雑音を重畳した音声データベースから雑音マッチドモデルを作成する．
2. データベースの話者ごとの十分統計量 (HMM における平均，分散，EM カウント) を 1. の雑音マッチドモデルから計算して，保存しておく．
3. 適応対象の雑音下の任意の一発声文に対して，GMM 話者モデルを用いてテスト話者に音響的特徴が近い話者を N 人選択する．
4. 3. で選択された N 人の話者について，2. の十分統計量を用いて適応モデルを再構築する．
5. 入力音声に対して SS を施した後に，30 dB SNR のオフィス雑音を重畳する．
6. 5. の音声を 4. の適応モデルで認識する．

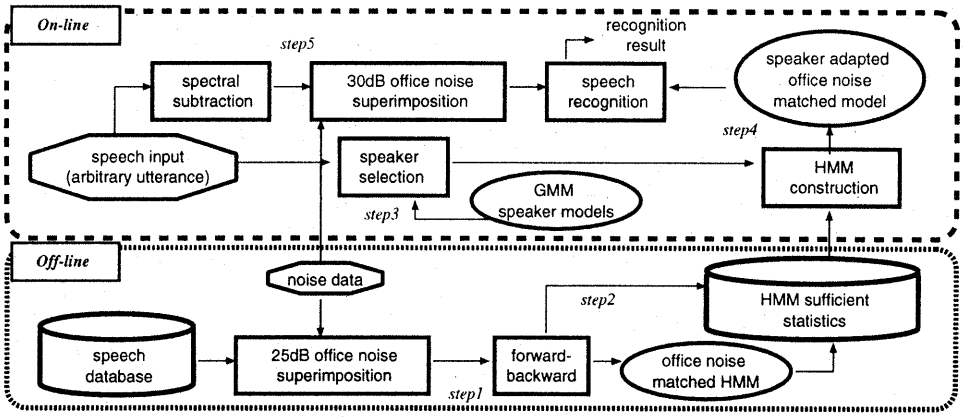

図 4　雑音に頑健な教師なし話者適応アルゴリズム

Fig. 4　Unsupervised speaker adaptation based on noise robust speech recognition

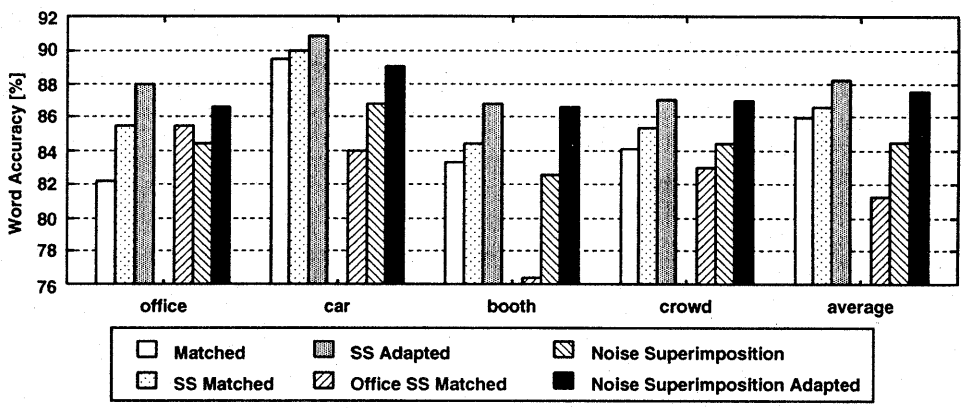

図 5　雑音に頑健なアルゴリズムに基づく教師なし話者適応の評価結果

Fig. 5　Word accuracy improvement by unsupervised speaker adaptation based on noise robust speech recognition

以上により，環境の変動および話者の交替に対応できる．

3.2 認識実験

提案した適応アルゴリズムの評価実験を行った．実験条件は，2.3.1 節の実験条件の表 3 と同一である．まず，オフラインでオフィス雑音を 25 dB SNR で重畳した学習用音声データベースから不特定話者オフィス環境マッチドモデルを作成し，そのマッチドモデルを用いて各話者ごとの十分統計量を計算し，保存しておく．また，学習用音声データベースにおける話者 GMM を 1 状態 64 混合で作成しておく．オンラインでは，任意の一発声文から，GMM 話者モデルを用いて，40 人の音響的特徴の近い話者を選択する [2].

入力音声の SNR を 20 dB としたときの認識実験の結果を図 5 に示す．また比較のために図 2 の結果を併せて示す．縦軸は単語正解精度 (%)，横軸は各環境雑音および実験条件を示している．Matched, SS Matched, Office SS Matched, Noise Superimposition は 2.3 節の実験結果である．SS Adapted は従来法 [3] にあたり，SS Matched に対して十分統計量により話者適応を行ったモデルでの認識率である．Noise Superimposition Adapted が提案した適応手法であり，Noise Superimposition に十分統計量による話者適応を行ったモデルでの認識率である．つまり Matched, SS Matched, SS Adapted が目的雑音ごとにマッチドモデルを用意した場合の結果，Office SS Matched, Noise Superimposition, Noise Superimposition Adapted が，目的雑音によらずオフィス環境のモデルを使用した場合の結果である．

話者適応を行った SS Adapted, Noise Superimposition Adapted は，各環境において高い適応の効果が得られた．特に booth においては高い適応の効果が得られ，Noise Superimposition で 82.6%であったものが，Noise Superimposition Adapted の 86.6%に改善された．office 以外の 3 種類の雑音を平均すると，84.6%から 87.5%へ 2.9%認識率が改善された．これは，環境ごとに十分統計量を作成して適応する SS Adapted と比較しても 0.7%の低下に抑えられた．

以上の結果から，提案する SS と雑音重畳に基づく教師な

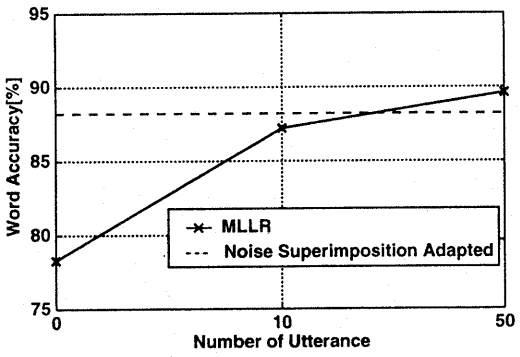

図 6　教師あり MLLR との比較評価結果
Fig. 6　Comparison with supervised MLLR adaptation

し環境・話者適応アルゴリズムは，対象雑音の種類ごとに SS Matched モデルを作成する従来法 [3] に近い認識性能を，雑音マッチドモデルを作成し直すことなく達成でき，一種類の十分統計量から高速に環境・話者に同時に適応可能であることが示せた．

3.3　教師あり MLLR との比較

現在最も一般的な教師ありの話者・環境適応手法は，MLLR [8] である．MLLR は適応データに対する尤度を最大化するように線形行列を求め，その行列を用いて HMM の平均，分散ベクトルを変換することにより音韻モデルを適応させるアルゴリズムである．MLLR はあらかじめ決められた発話内容を確実に読み上げることで，高精度な適応モデルが得られるとされている．そこで，本提案手法と教師あり MLLR との比較を行う．

MLLR は不特定話者 clean モデルを初期モデルとして平均と分散に対し，教師あり 10 文章および 50 文章を用い，3 回の繰り返し学習を行った．car，booth，crowd の 3 種類環境に対する MLLR での適応結果の平均認識率を図 6 に，詳細な結果は表 4 に示す．比較のために提案手法である SS と雑音重畳に基づく教師なし環境・話者適応法における平均の結果を併せて示す．提案法の結果は，適応に用いるデータが一文章のみであるので直線で表す．横軸は，適応に用いた文章数，縦軸は単語正解精度 (%) である．図 6 より，教師あり MLLR の 10 文章より高い認識率を示し．MLLR 適応は数十文章のユーザ発声が要求され，ユーザに大きな負担をかけるが，提案手法は，適応に用いるデータは任意の一発声文章のみで良いため，ユーザにかかる負担はかなり軽減される．

表 4　教師あり MLLR の結果
Table 4　Experiment result of supervised MLLR adaptation

	Number of sentences		
noise type	before	10 utterances	50 utterances
office	65.2	82.2	85.6
car	85.2	90.9	91.7
booth	71.8	84.2	88.0
crowd	78.0	86.7	89.2
average	78.3	87.2	89.6

4.　おわりに

本稿では，SS と雑音重畳に基づく雑音に頑健な音声認識アルゴリズムを提案した．また，提案アルゴリズムを十分統計量を用いた教師なし話者適応アルゴリズムと統合することにより，環境雑音の種類や SNR の変動に頑健で，かつ任意の一発声文のみを用いて高速に話者に適応できる．認識実験より，20 dB SNR の環境で平均約 88%の認識精度が得られた．最後に，現在最も一般的な適応法である教師あり MLLR 法と比較を行い，10 文章を用いた MLLR による適応モデルよりも，提案手法による適応モデルの方が認識性能が高いことを示した．

謝辞　本研究の一部は科学技術振興事業団による戦略的研究推進事業 (CREST)「高度メディア社会の生活情報技術」の援助を受けて行われた．

文　献

[1] S. Yoshizawa, A. Baba, K. Matsunami, Y. Mera, M. Yamada, K. Shikano: "Unsupervised Speaker Adaptation Based on Sufficient HMM Statistics of Selected Speakers," Proceedings of ICASSP, pp.341–344, 2001.

[2] S. Yoshizawa, A. Baba, K. Matsunami, Y. Mera, M. Yamada, K. Shikano: "Evaluation on Unsupervised Speaker Adaptation Based on Sufficient HMM Statistics of Selected Speakers," Proceedings of EuroSpeech, pp.1219–1222, 2001.

[3] S. Yamade, K. Matsunami, A. Baba, A. Lee, H. Saruwatari, K. Shikano: "Spectral Subtraction in Noisy Environments Applied to Speaker Adaptation Based on HMM Sufficient Statistics," Proceedings of ICSLP2002, pp.1045–1048, 2002.

[4] S. F. Boll: "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," IEEE Transaction on Acoustics Speech and Signal Processing, ASSP–33, vol.27, pp.113–120, 1979.

[5] 電子協騒音データベース，http://it.jeita.or.jp/jhistory/committee/humanmed/speech/noisedbj.html

[6] K. Itou, M. Yamamoto, K. Takeda, T. Takezawa, T. Matsuoka, T. Kobayashi, K. Shikano, S. Itahashi: "JNAS : Japanese Speech Corpus for Large Vocabulary Continuous Speech Recognition Research," The Journal of the Acoustical Society of Japan(E), vol.20, pp.199–206, 1999.

[7] T. Kawahara, et al.: "Free Software Toolkit for Japanese Large Vocabulary Continuous Speech Recognition," Proceedings of ICSLP, Ob(16)-V-07, pp.IV-476–479, 2000.

[8] C. J. Leggetter, C. Woodland: "Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density Hidden Markov Models," Computer Speech and Language, vol.9, pp.701–704, 1995.

[9] A. Lee, T. Kawahara, K. Takeda, K. Shikano: "A New Phonetic Tied Mixture Model for Efficient Decoding," Proceedings of ICASSP, pp.1269–1272, 2000.

[10] M. Yamada, A. Baba, S. Yoshizawa, Y. Mera, A. Lee, H. Saruwatari, K. Shikano: "Unsupervised Noisy Environment Adaptation Algorithm Using MLLR and Speaker Selection," Proceedings of EuroSpeech, pp.869–872, 2001.