修 士 論 文

ヤコビ適応法を用いた雑音・伝達特性・発声変形への同時適応

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

坂井 伸圭

2002年3月

修 士 論 文

ヤコビ適応法を用いた雑音・伝達特性・発声変形への同時適応

指導教官　嵯峨山 茂樹 教授

審査委員主査　嵯峨山 茂樹 教授
審査委員　下平 博 助教授
審査委員　赤木 正人 教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

010047坂井 伸圭

提出年月: 2002年2月

概 要

高騒音環境下における音声認識での認識率低下の原因としては，背景雑音やマイクロフォンの伝達特性に加えて音声そのものの発声変形（Lombard 効果）がある．これらの要因に対して同時に高速に適応する必要がある．ヤコビ適応法は，Taylor 展開の 1 次項を用いて近傍を線形近似し，計算量を削減した高速な適応法である．本研究では，ヤコビ適応法を用いた発声変形への適応法を提案し，さらに雑音・伝達特性・発声変形への同時適応を行う手法について提案し，その効果を実験を通して確認する．

目 次

図目次

表目次

第1章　序論

1.1　研究の背景と目的

現在，音声認識技術は，統計的手法の発達により，雑音の無い理想的な環境においては高い認識性能を示す．しかし，雑音や伝達特性が存在する実環境においては，認識率が大きく低下する．これは，作成した音響モデルと発話環境での音響モデルとの間にミスマッチが生じる為である．ミスマッチの原因としては，雑音や伝達特性等の音声に重畳される背景雑音による問題と，背景雑音の存在によって発話者がより聞こえやすいように発話しようとするために起こる発声変形 (Lombard 効果)[1] の問題とがある．これらの変動要因に対して，音響モデルを適応させていく必要がある．これら個々の要因に対する適応法としては従来様々な研究がなされてきた．しかし，これらの要因に対して同時に適応する手法に関しては複雑な非線型問題となるため，あまり研究されていない．そこで，複雑な非線型問題を，Taylor 展開の 1 次項を用いて近傍を線形近似して解く，という特徴を持つヤコビ適応法 [2]-[7] を用いて，これらの要因に対して音響モデルの同時適応を行う手法を検討する．既に，ヤコビ適応法を用いた雑音・伝達特性に対する適応については，報告されている [8]．また，前年度の研究により，話者に対してのヤコビ適応法が提案された．

本研究では，まず，前年度までの研究を統合し，雑音・伝達特性・話者に対する同時適応法を提案し，実験を通してその効果を確認する．さらに話者に対するヤコビ適応の手法を拡張した発声変形に対するヤコビ適応法について提案し，ヤコビ適応法を用いた雑音・伝達特性・発声変形に対する同時適応を行い，実験を通してその効果を確認する．

1.2　本論文の構成

第 1 章では, 序論として既に研究の背景と目的について述べた. 第 2 章では, 嵯峨山らによって提案されたヤコビ適応法の原理について簡単に述べる. 第 3 章では, ヤコビ適応法を用いた雑音・伝達特性・話者への同時適応について，その手法と同時適応システムの説明，さらに実験結果について述べる．第 4 章では, 発声変形によるスペクトル変化の特徴，そして，ヤコビ適応を用いた発声変形への適応手法について述べる．第 5 章では, ヤコビ適応法を用いた雑音・伝達特性・発声変形に対する同時適応の手法について述べる．最後に第 5 章は結論と今後の課題とし, 本研究で得られた結果とその考察，また今後の課題について述べる．

第2章　ヤコビ適応法の基本原理

本章では，まずヤコビ適応法の基本原理について説明し，次に加法性歪みである雑音に対するヤコビ適応法についてそのアルゴリズムと処理の流れについて説明する．

2.1　ヤコビ適応法

ヤコビ適応法は，実時間処理に適した音響モデル適応法として提案されている [2]-[7]. ヤコビ適応法の特徴としては，加法性歪みである雑音成分をケプストラム領域において重畳する際, モデルパラメータの更新に必要となる非線形演算を，Taylor 展開の 1 次項を用いた線形近似式で演算することにより高速なモデルの適応を行うという特徴がある．この特徴により，適応に必要なデータを観測してから音響モデルを適応するまでの処理時間が少なく，また，少量の観測データでも適応が可能とされている．そのため，実時間処理に向いた雑音耐性技術として近年注目を浴びている．

2.1.1　基本原理

あるスペクトルが n 次元ベクトル $X = (x_1, x_2, \ldots, x_N)^T$ と，$Y = (y_1, y_2, \ldots, y_N)^T$ のどちらでも表現できる場合，これらの両領域での微小変化 $\Delta X, \Delta Y$ には式 (2.1) の関係がある．

$$\Delta Y = \frac{\partial Y}{\partial X} \Delta X \tag{2.1}$$

ここで，$\Delta X = (\Delta x_1, \Delta x_2, \ldots, \Delta x_N)^T$, $\Delta Y = (\Delta y_1, \Delta y_2, \ldots, \Delta y_N)^T$ で，$[\frac{\partial Y}{\partial X}]_{ij} = \frac{\partial y_i}{\partial x_j}$ は X と Y の間のヤコビ行列と呼ばれる． これは，X の領域にて微小変動が観測されれば，Y の領域での修正項を計算できることを示している．これが，以降の雑音適応，話者適応，発声変形適応に成り立つ基本原理となる．次節では,X を雑音のケプストラム，Y を雑音重畳音声のケプストラムとして雑音適応の場合について論じる．

2.1.2 ヤコビ適応法による雑音適応

雑音重畳音声の線形スペクトル S_Y(ベクトル表現である)は，クリーン音声の線形スペクトル S_S と雑音の線形スペクトル S_N の加算によって表される．

$$S_Y = S_S + S_N \tag{2.2}$$

また，線形スペクトルから対数スペクトルへは対数をつかい変換を行う．そこで，線形スペクトル領域からケプストラム領域へと変換した場合のクリーン音声ケプストラム C_S，雑音ケプストラム C_N，雑音重畳音声ケプストラム C_Y の関係は下式で表される．

$$C_Y = F^{-1}[\log\{\exp(FC_S) + \exp(FC_N)\}] \tag{2.3}$$

ここで，F,F^{-1} はそれぞれフーリエ変換，逆フーリエ変換であり，exp, log はそれぞれ指数変換，対数変換である．一般に，音響モデルの特徴量としてはケプストラムが用いられる．そのため，雑音適応を音響モデルを用いて行う場合は式 (2.3) の演算を処理する必要があり，その計算量は大きなものとなってしまう．

そこでヤコビ適応法においては，雑音の変化が微小であると考え，式 (2.4) で示される Taylor 展開の 1 次微分項を利用して計算量を減少している．

$$f(x+\Delta x) = f(x) + \frac{f'(x)}{1!}\Delta x + \frac{f''(x)}{2!}(\Delta x)^2 + \cdots + \frac{f^{n-1}(x)}{n-1!}(\Delta x)^{n-1} + \frac{f^n(x+\theta\Delta x)}{n!}(\Delta x)^n \tag{2.4}$$

具体的には，雑音重畳音声，音声，雑音の変動分を考えた場合，下式のように求まる．

$$\Delta C_Y \simeq \frac{\partial C_Y}{\partial C_S}\Delta C_S + \frac{\partial C_Y}{\partial C_N}\Delta C_N \tag{2.5}$$

式 (2.5) に表されているように，変動分の演算が線形近似されているために，ヤコビ適応法においては線形スペクトル領域への変換を必要としない． そのため，従来の手法に比べて高速な適応が可能となる． 図 2.1 にヤコビ適応法と HMM 合成法との比較を，図 2.2 にヤコビ適応の概念図を示す．

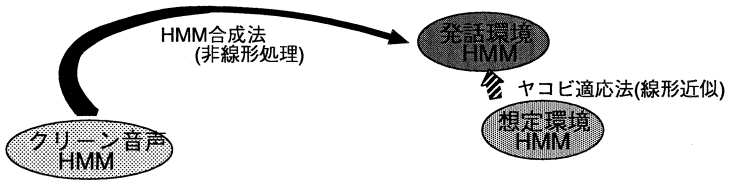

図 2.1: ヤコビ適応法と HMM 合成法との比較

そこでヤコビ適応法においては，事前に想定した環境 (環境 A) とその環境のヤコビ行列を予め用意しておき，発話者周辺の環境 (環境 B) との差分を用いて音響モデルを修正

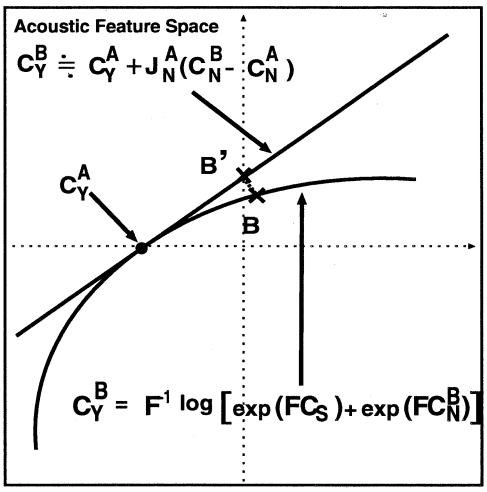

図 2.2: ヤコビ適応の概念図

する流れで適応が行われる．環境 A の雑音ケプストラムを C_N^A，雑音重畳音声ケプストラム C_Y^A，ヤコビ行列を J_N とし，また，環境 B の雑音ケプストラムを C_N^B 雑音重畳音声ケプストラム C_Y^B とすると，式 (2.6) と表される．

$$C_Y^B \simeq C_Y^A + J_N(C_N^B - C_N^A) \tag{2.6}$$

なお，$\frac{\partial C_Y}{\partial C_N} = J_N$ であり，式 (2.5) における ΔC_S を 0，つまり，ロンバート効果等による発話者の発話様式が変化しないと仮定する．

次に，実際にヤコビ行列 ($\frac{\partial C_Y}{\partial C_N}$) は，式 (2.7) のように展開することで求められる．

$$J_N \equiv \frac{\partial C_Y}{\partial C_N} = \frac{\partial C_{S+N}}{\partial \log S_{S+N}} \frac{\partial \log S_{S+N}}{\partial S_{S+N}} \frac{\partial S_{S+N}}{\partial S_N} \frac{\partial S_N}{\partial \log S_N} \frac{\partial \log S_N}{\partial C_N} = F^{-1} \frac{S_N}{S_{S+N}} F \tag{2.7}$$

なお，それぞれの項は式 (2.8)〜(2.12) に示されるように計算できる ($\mathrm{F}, \mathrm{F}^{-1}, \mathrm{E}$ をそれぞれコサイン変換行列，逆コサイン変換行列，単位行列とする)．

- $C_{Yi} = \mathrm{F}_{ij}^{-1} \log S_{S+N_j}$ より

$$\left[\frac{\partial C_Y}{\partial \log S_{S+N}}\right]_{ij} = \mathrm{F}^{-1}{}_{ij} \tag{2.8}$$

- $(\log S_{S+N_i}) = \log(S_{S+N_i})$ より

$$\left[\frac{\partial \log S_{S+N}}{\partial S_{S+N}}\right]_{ij} = \mathrm{E}_{ij} \frac{1}{S_{S+N\,i}} \tag{2.9}$$

- $S_{S+N_i} = S_{Si} + S_{Ni}$ より

$$\left[\frac{\partial S_{S+N}}{\partial S_N}\right]_{ij} = \mathrm{E}_{ij} \tag{2.10}$$

- $S_{Ni} = exp(\log S_{Ni})$ より

$$\left[\frac{\partial S_N}{\partial \log S_N}\right]_{ij} = \mathrm{E}_{ij} S_{Ni} \tag{2.11}$$

- $\log S_{Ni} = \mathrm{F}_{ij} C_{Ni}$ より

$$\left[\frac{\partial \log S_N}{\partial C_N}\right]_{ij} = \mathrm{F}_{ij} \tag{2.12}$$

実際にヤコビ行列の各要素を求めるには，クリーン音声の音響モデルの平均ベクトルの線形スペクトルを S_S，雑音重畳音声音響モデルの平均ベクトルの線形スペクトルを S_N，F, F^{-1} をそれぞれコサイン変換行列，逆コサイン変換行列とすると，式 (2.13) より求められる．

$$[J_N]_{ij} = \sum_k F_{ik}^{-1} \frac{S_{N_k}}{S_{S_k} + S_{N_k}} F_{kj} \tag{2.13}$$

なお，平均ベクトルに対する適応は式 (2.14)，共分散行列に対する適応は (2.15) となる．

- 平均パラメータの更新式

$$Mean[C_Y^B] \simeq Mean[C_Y^A] + J_N \Delta Mean[C_N] \tag{2.14}$$

- 分散パラメータの更新式

$$\begin{aligned} Cov[C_Y^B] &\simeq Cov[C_Y^A + \Delta C_Y] \\ &\simeq Cov[C_Y^A + J_N \Delta C_N] \\ &\simeq Cov[C_Y^A] + J_N \Delta Cov[C_N] J_N^T \end{aligned} \tag{2.15}$$

2.1.3 適応アルゴリズム

実際の適応処理の流れは図 2.3 に示される通りで，その処理は以下の事前処理・適応処理・認識処理の 3 段階に分けられる．

- 事前処理
  1. 環境 A の雑音から初期雑音 HMM(C_N^A) を求める
  2. 環境 A を用いて雑音重畳音声 HMM(C_Y^A) を求める
  3. ヤコビ行列 (J_N) を計算する

- 適応処理
  1. 環境 B で発声直前に観測される雑音から，観測雑音 HMM(C_N^B) を求める
  2. 初期雑音 HMM と観測雑音 HMM から，雑音の変化量 (ΔC_N^B) を求める

3. ヤコビ行列と雑音の変化量から，雑音重畳音声の変化量 (ΔC_Y^B) を求める
4. 雑音重畳音声の変化量から初期モデルを修正し，環境Bの雑音重畳音声HMM(C_Y^B)を求める

- 認識処理
  1. 適応したHMMモデルを用いて，環境Bの音声に対する認識を行う

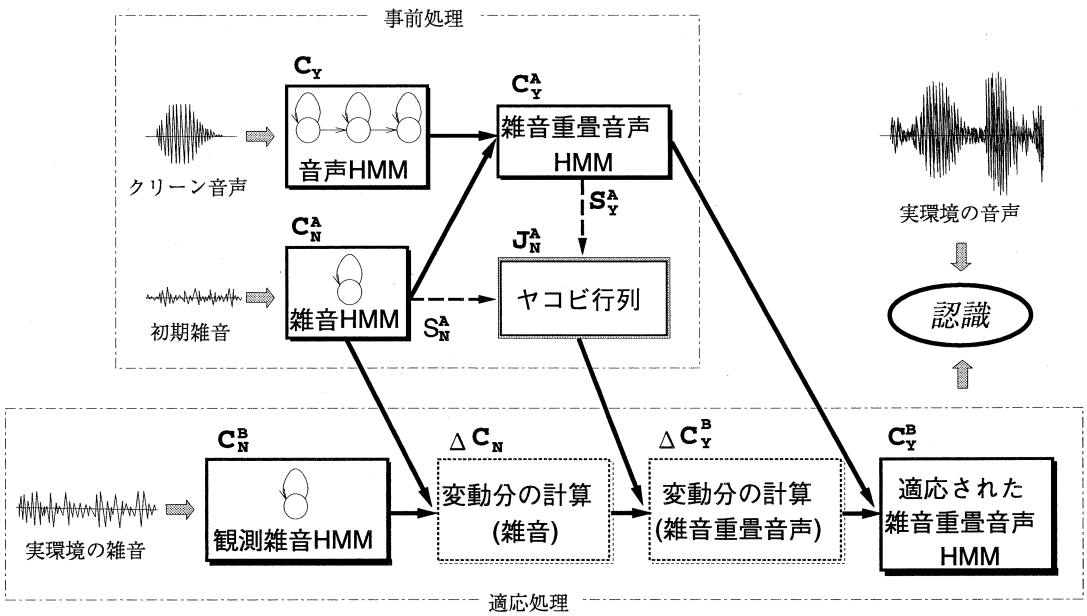

図 2.3: 雑音環境に対するヤコビ適応の処理の流れ

第3章　ヤコビ適応法を用いた雑音・伝達特性・話者への同時適応

音響モデルの学習環境と認識環境のミスマッチによる認識率低下の要因として，背景雑音，マイクロフォン等の伝達特性，および話者性等が挙げられる．これらの要因が時々刻々と変化する実環境においては，モデルを高速に同時適応する必要がある．先行研究 [8] では，雑音・伝達特性への適応を同時に行う手法を報告されているが，本章では，話者へのヤコビ適応法を示した後，雑音・伝達特性・話者への同時適応法を示し，実験によってその効果を示す．

3.1　ヤコビ適応法による話者適応

3.1.1　基本原理

ここでは，話者性のうち，声道長の相違に対して適応する方法を考える．声道長による相違は，音声スペクトル $S(\omega)$ に対する，周波数軸の伸縮係数 λ ($\lambda \approx 1$) による線形伸縮スペクトル $S(\lambda\omega)$ としてモデル化できる (図 3.1)．この場合，$\lambda < 1$ は周波数軸の伸長，$\lambda > 1$ は収縮を意味する．

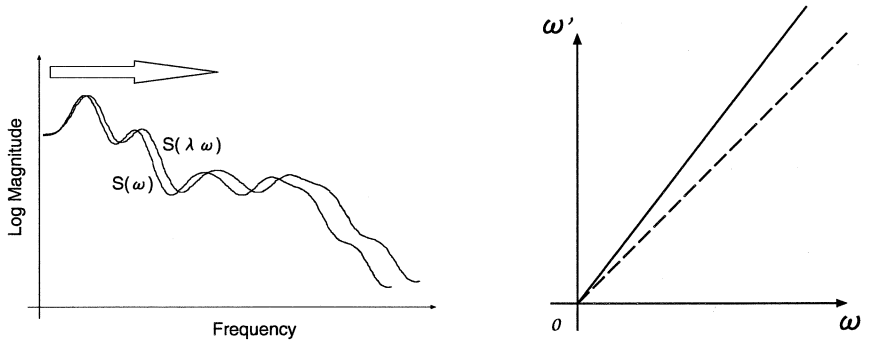

図 3.1: スペクトルの変化と ω の変化

音声スペクトル $S(\lambda\omega)$ のケプストラム (p 次元ベクトル) を $\boldsymbol{C}_\lambda = \{C_{\lambda i}\}_1^p$ で記すと，$\boldsymbol{C}_\lambda$ と伸縮前の元のケプストラム $\boldsymbol{C}$ との関係は次式で表せる．

$$\boldsymbol{C}_\lambda = \boldsymbol{F}^{-1}\boldsymbol{F}_\lambda\boldsymbol{C} \tag{3.1}$$

ここで，$\boldsymbol{F}^{-1}$ は，離散コサイン変換行列 $(\boldsymbol{F})$ の逆行列，$\boldsymbol{F}_\lambda$ は周波数伸縮を伴う離散コサイン変換行列で，その (i,j) 成分は次式で与えられる．

$$(\boldsymbol{F}_\lambda)_{ij} = \cos\left(\frac{\lambda i(2j+1)\pi}{2N}\right) \tag{3.2}$$

ここで，N は周波数の標本化数である．

今，声道長伸縮係数 λ が微少量 $\Delta\lambda$ だけ変化したとすると，対応する $\boldsymbol{C}_\lambda$ の変動量 $\Delta\boldsymbol{C}_\lambda$ は，Taylor 展開による線形近似によって次式で与えられる．

$$\Delta\boldsymbol{C}_\lambda = \frac{\partial\boldsymbol{C}_\lambda}{\partial\lambda}\Delta\lambda \tag{3.3}$$

ここで，$\boldsymbol{J}_\lambda = \partial\boldsymbol{C}_\lambda/\partial\lambda$ と置き，これを声道長ヤコビ行列と呼ぶことにする．$\lambda \approx 1$ のとき，$\boldsymbol{J}_\lambda$ は (3.1) 式より，

$$\boldsymbol{J}_\lambda = \boldsymbol{F}^{-1}\frac{\partial F^{(\lambda)}}{\partial\lambda}\boldsymbol{C} \;\approx\; \boldsymbol{F}^{-1}\boldsymbol{G}\boldsymbol{C} \tag{3.4}$$

で与えられる．ここで，

$$(\boldsymbol{G})_{ij} = \frac{-i(2j+1)\pi}{2N}\sin\left(\frac{i(2j+1)\pi}{2N}\right) \tag{3.5}$$

であり，$\boldsymbol{J}_\lambda$ は初期モデル作成時に求めておくことができる．

3.2 ヤコビ適応を用いた雑音・伝達特性・話者への同時適応

3.2.1 基本原理

本稿では図 3.2 に示す観測系を想定する．初めに仮定した初期モデルから発話環境に適応したモデルへのケプストラムの変動分 $\Delta\boldsymbol{C}_Y$ は，声道長ヤコビ行列 $\boldsymbol{J}_\lambda$，雑音ヤコビ行列 $\boldsymbol{J}_N$ を用いて，次式で表せる．

$$\Delta\boldsymbol{C}_Y = \boldsymbol{J}_\lambda\Delta\lambda + \boldsymbol{J}_N\Delta\boldsymbol{C}_N + \Delta\boldsymbol{C}_H \tag{3.6}$$

$\Delta\lambda$，$\Delta\boldsymbol{C}_N$，$\Delta\boldsymbol{C}_H$ はそれぞれ声道長伸縮係数の変動分，雑音の変動分，伝達特性の変動分を表す．ここで $\boldsymbol{J}_N$ は次式で与えられる．

$$\boldsymbol{J}_N \equiv \frac{\partial\boldsymbol{C}_Y}{\partial\boldsymbol{C}_N} = \boldsymbol{F}^{-1}\frac{\boldsymbol{S}_N}{\boldsymbol{S}_S + \boldsymbol{S}_N}\boldsymbol{F} \tag{3.7}$$

観測した適応用音声には，話者変動成分，雑音変動分，および伝達特性変動分が含まれているため，1つの代表点(例えば，HMMの1つの出力分布の平均ベクトル)から，それぞれの変動成分を分離して求めることはできない．そこで，最小自乗誤差法を用いて多数の代表点から (3.6) 式の $\Delta \boldsymbol{C}_N$ ，$\Delta \boldsymbol{C}_H$，$\Delta\lambda$ を推定する．

音素 HMM の状態 i における平均ベクトルの初期環境と観測環境の変動分を $\Delta\widehat{\boldsymbol{C}_Y}^{(i)}$，観測に伴う誤差ベクトルを $\boldsymbol{\epsilon}^{(i)}$ とし，全ての音素モデルの総状態数を M とすると，推定すべき $\Delta\widehat{\lambda}$，$\Delta\widehat{\boldsymbol{C}_N}$，$\Delta\widehat{\boldsymbol{C}_H}$ は以下の M 個の連立方程式における誤差ベクトルの自乗和を最小化する事によって得られる．

$$
\begin{cases}
\Delta\widehat{\boldsymbol{C}_Y}^{(1)} &= \boldsymbol{J}_\lambda^{(1)}\Delta\widehat{\lambda} + \boldsymbol{J}_N^{(1)}\Delta\widehat{\boldsymbol{C}_N} + \Delta\widehat{\boldsymbol{C}_H} + \boldsymbol{\epsilon}^{(1)} \\
\Delta\widehat{\boldsymbol{C}_Y}^{(2)} &= \boldsymbol{J}_\lambda^{(2)}\Delta\widehat{\lambda} + \boldsymbol{J}_N^{(2)}\Delta\widehat{\boldsymbol{C}_N} + \Delta\widehat{\boldsymbol{C}_H} + \boldsymbol{\epsilon}^{(2)} \\
&\vdots \\
\Delta\widehat{\boldsymbol{C}_Y}^{(M)} &= \boldsymbol{J}_\lambda^{(M)}\Delta\widehat{\lambda} + \boldsymbol{J}_N^{(M)}\Delta\widehat{\boldsymbol{C}_N} + \Delta\widehat{\boldsymbol{C}_H} + \boldsymbol{\epsilon}^{(M)}
\end{cases}
\tag{3.8}
$$

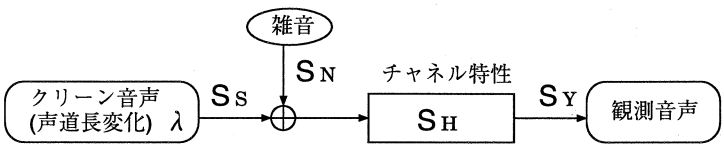

図 3.2: 仮定する音声信号の観測系

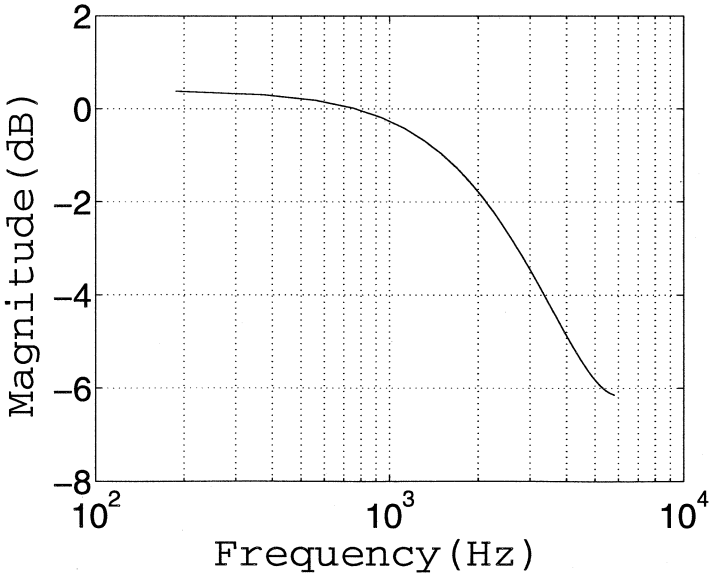

図 3.3: 観測雑音環境における伝達特性

3.2.2 適応アルゴリズム

実際の適応処理の流れは図3.4に示される通りで，その処理は以下の事前処理・適応処理・認識処理の3段階に分けられる．

事前処理

1. 事前に想定した雑音環境Aに基づき PMC 法により雑音重畳音声 HMM ($\widehat{\boldsymbol{C}_Y}^A$) を作成 → 初期モデル A
2. ヤコビ行列の計算： → $\boldsymbol{J}_N$,$\boldsymbol{J}_\lambda$

適応処理

1. 観測環境 B における音声サンプルより HMM ($\widehat{\boldsymbol{C}_Y}^B$) を作成（Viterbi time alignment を利用）→ 観測モデル B
2. 初期モデル A と観測モデル B の平均ベクトルの変化量 ($\Delta\widehat{\boldsymbol{C}_Y}$) を計算
3. $\Delta\widehat{\boldsymbol{C}_Y}$，$\boldsymbol{J}_N$，$\boldsymbol{J}_\lambda$ を用いて，最小自乗法により $\Delta\widehat{\boldsymbol{C}_N}$，$\Delta\widehat{\boldsymbol{C}_H}$，$\Delta\widehat{\lambda}$ を推定
4. 初期モデル A のパラメータを修正し，適応モデル B を作成

認識処理

1. 適応した適応モデル B を用いて，観測環境 の音声に対する認識を行う

3.3 評価実験

3.3.1 実験条件

表3.1の通り，条件 (I) と条件 (II) の 2 通りの実験を行った．初期環境の音響モデルは，クリーン音声で学習した環境独立型音素 HMM と初期雑音で学習した 1 状態 1 分布の雑音 HMM を用いて HMM 合成法により作成した．雑音の観測時間は 60 秒である．観測音声としては，初期環境とは異なる話者のクリーンな単語音声に異なる雑音と伝達特性を重畳したものを使用した．

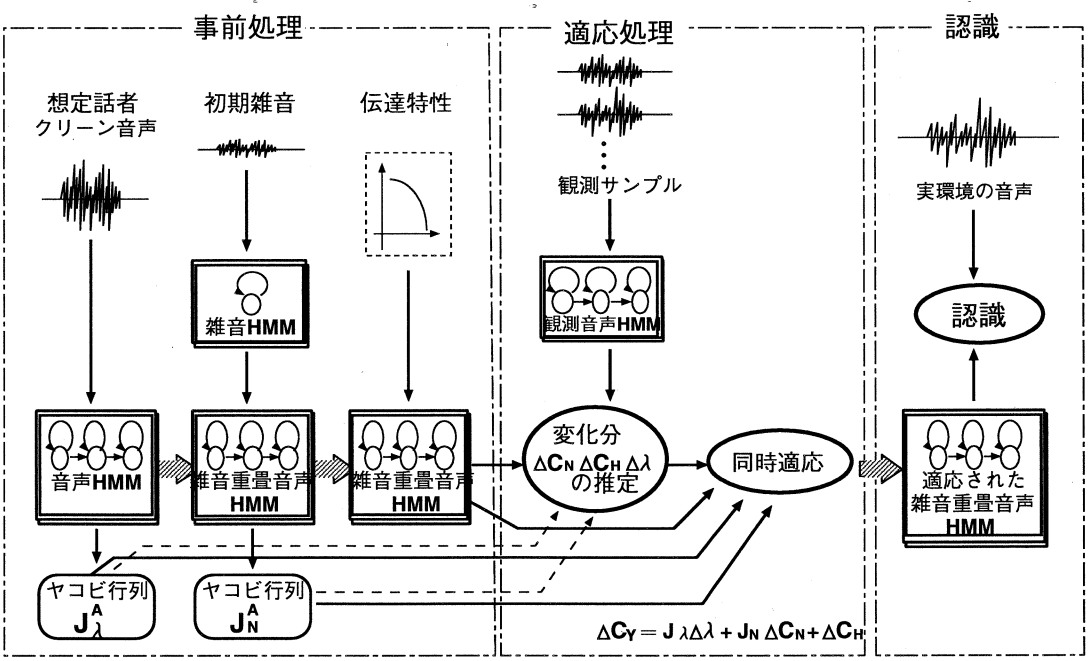

図 3.4: 雑音・伝達特性・話者へのヤコビ同時適応システム処理の流れ

表 3.1: 実験条件

音声 DB	ATR 音声データベース 重要単語 5240 単語
雑音 DB	電子協騒音データベース
学習資料	2620 単語（奇数番）
評価資料	655 単語（偶数番の半分）
特徴量	16 次 LPC + 16 次 Δ LPC（0 次を含む）
音素 HMM	3 状態 3 混合，環境独立型
条件 (I)：男性話者間の適応	
初期環境	観測環境
背景雑音：駅構内 伝達特性：全域通過型 話者　　：MAU（男性）	背景雑音：交差点 伝達特性：低域通過（図 3.3） 話者　　：MHT（男性）
条件 (II)：女性話者の適応	
初期環境	観測環境
背景雑音：車内 伝達特性：全域通過型 話者　　：FFS（女性）	背景雑音：工場 伝達特性：低域通過（図 3.3） 話者　　：FMS（女性）

3.3.2 適応の様子

実際にHMM合成法(PMC法)によって作成した初期モデルから，雑音・伝達特性・話者への同時適応により，どのようにスペクトルが変化するか，その変化の様子を図3.5～3.12に示す．図中のPMCは初期モデルのスペクトルを表し，adapt(8words),adapt(8words)は，初期モデルから適応単語数8words,64wordsにおける同時適応後のスペクトルを表し，closeは観測環境の音声によって学習した理想モデルのスペクトルになる．図3.5～3.12により，適応単語数を増やすにつれて，初期モデルから，観測環境モデルに近づいている様子が確認できた．

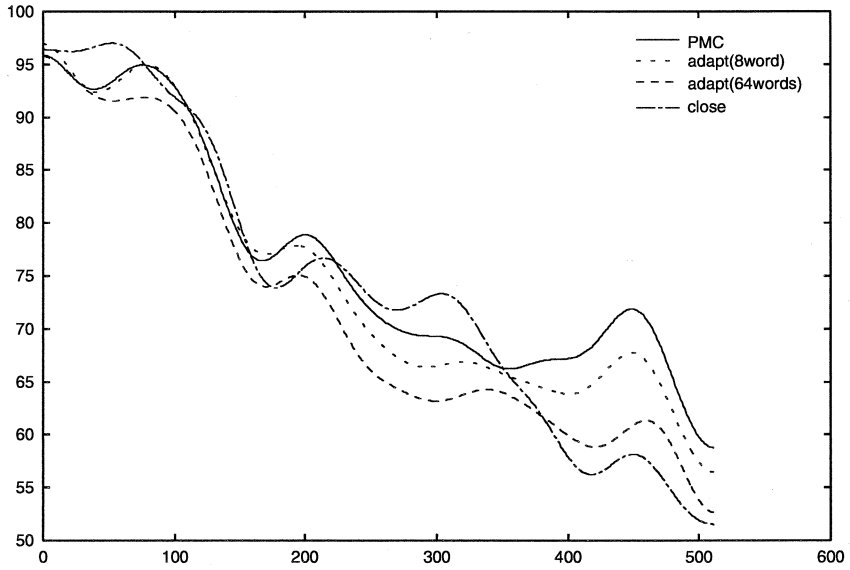

図 3.5: 音素 /a/ 話者 mau - mht 雑音 05 - 09 CNS 同時適応の様子

3.3.3 $\Delta\lambda$ の推定精度

条件(I)の適応で推定された $\Delta\lambda$ の値を表3.2に示す．また，$\Delta\lambda$ の最適値を実験的に調べるため，クリーン音声に対して λ を変化させた時の認識率を図3.13に示す．これらの結果より，適応単語数が多くなるにつれて，最適値に近い $\Delta\lambda$ の値 $(-0.01 \sim -0.03)$ が推定できていることが分かる．

3.3.4 認識結果

認識実験の結果を図3.14 ～ 3.16に示す．図3.14，図3.15より，話者性を含めて同時適応を行う提案手法の方が，話者適応を行わずに雑音・伝達特性にのみ同時適応する場合より認識率が向上していることが分かる．ただし，認識率の向上は男性話者(条件(I))で，1

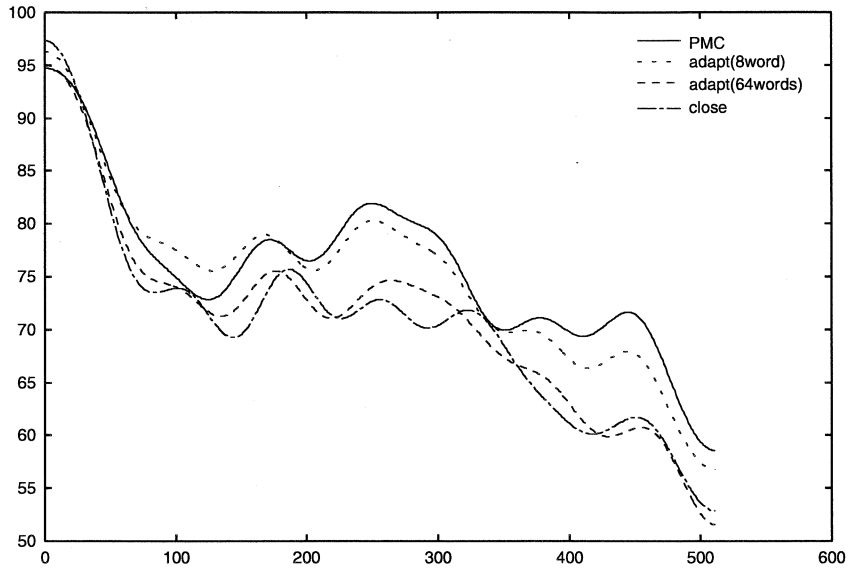

図 3.6: 音素 /i/ 話者 mau - mht 雑音 05 - 09 CNS 同時適応の様子

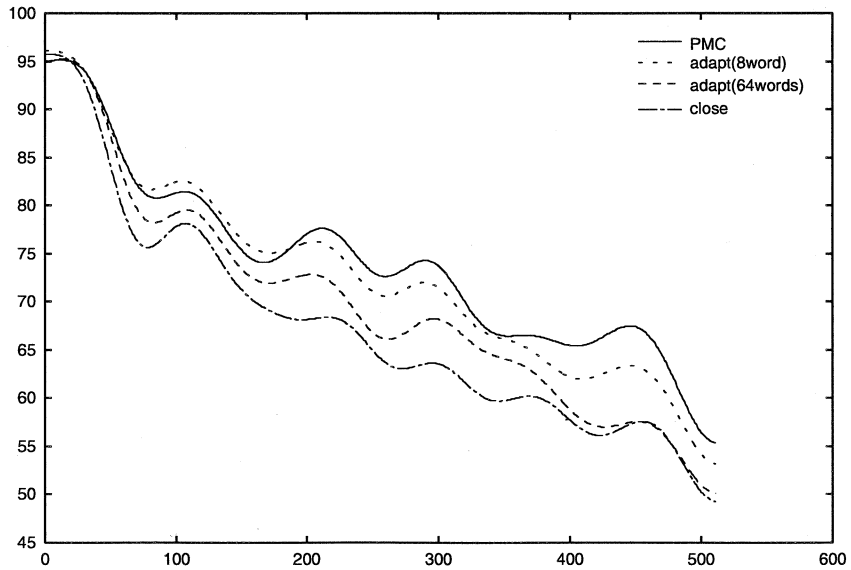

図 3.7: 音素 /u/ 話者 mau - mht 雑音 05 - 09 CNS 同時適応の様子

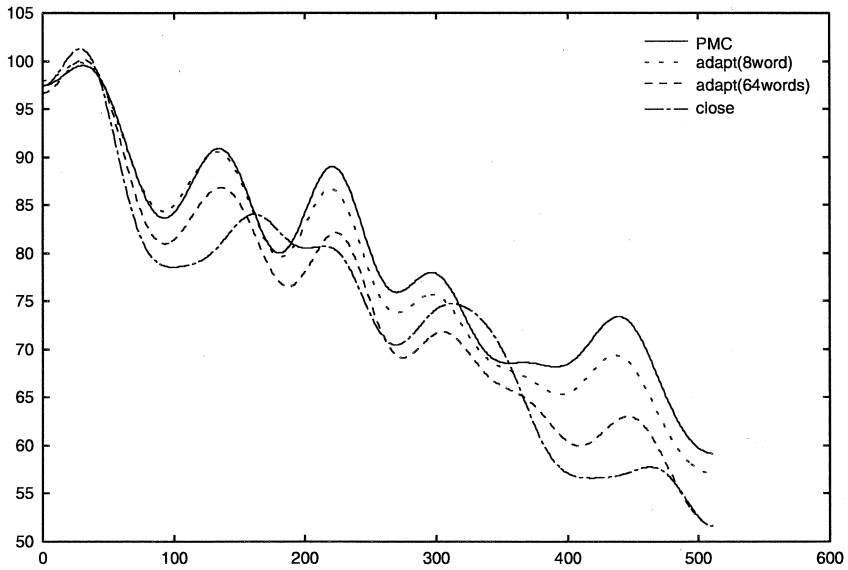

図 3.8: 音素 /e/ 話者 mau - mht 雑音 05 - 09 CNS 同時適応の様子

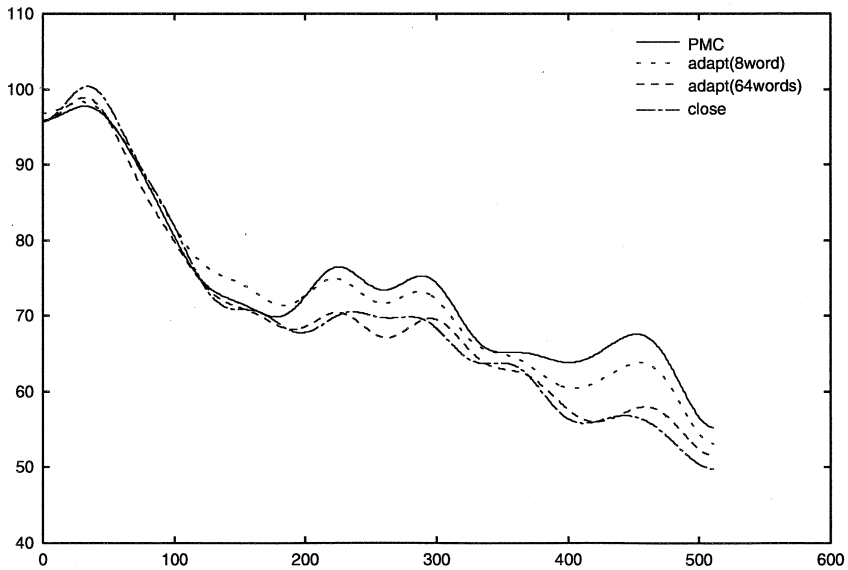

図 3.9: 音素 /o/ 話者 mau - mht 雑音 05 - 09 CNS 同時適応の様子

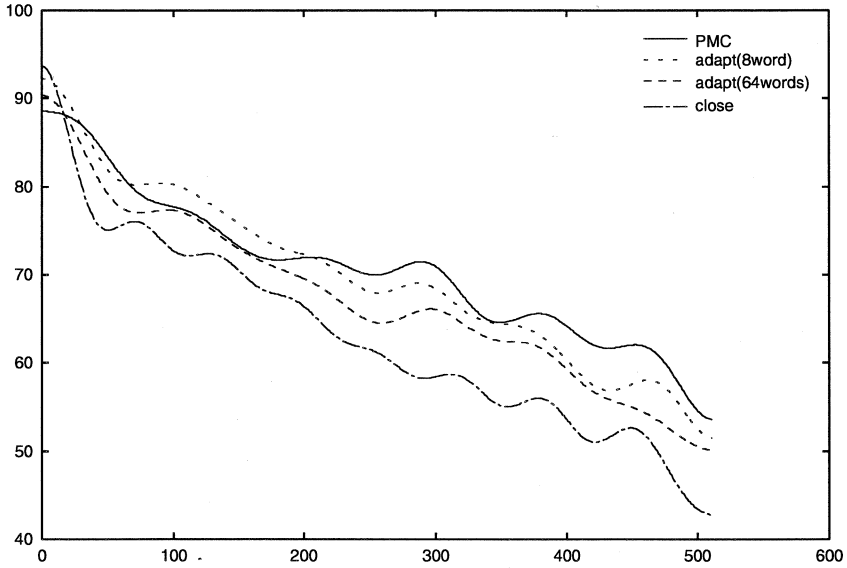

図 3.10: 音素 /p/ 話者 mau - mht 雑音 05 - 09 CNS 同時適応の様子

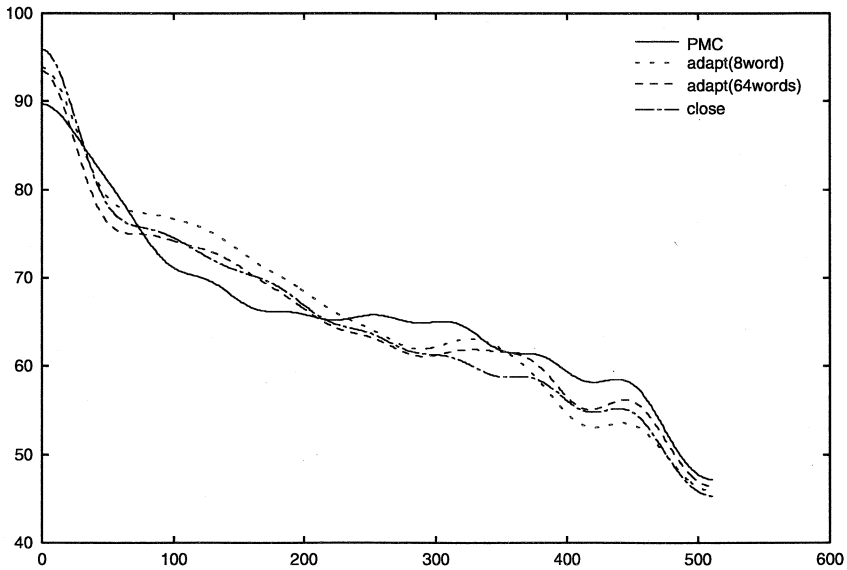

図 3.11: 音素 /b/ 話者 mau - mht 雑音 05 - 09 CNS 同時適応の様子

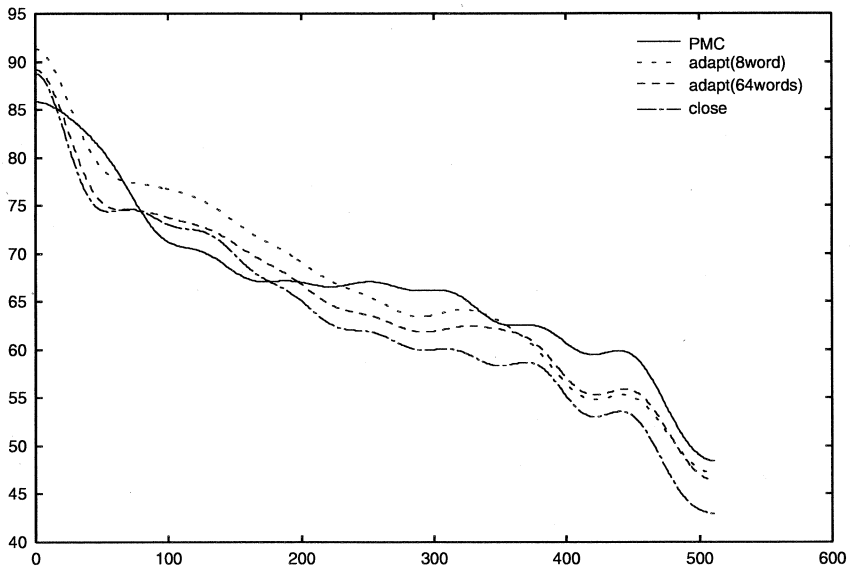

図 3.12: 音素 /k/ 話者 mau - mht 雑音 05 - 09 CNS 同時適応の様子

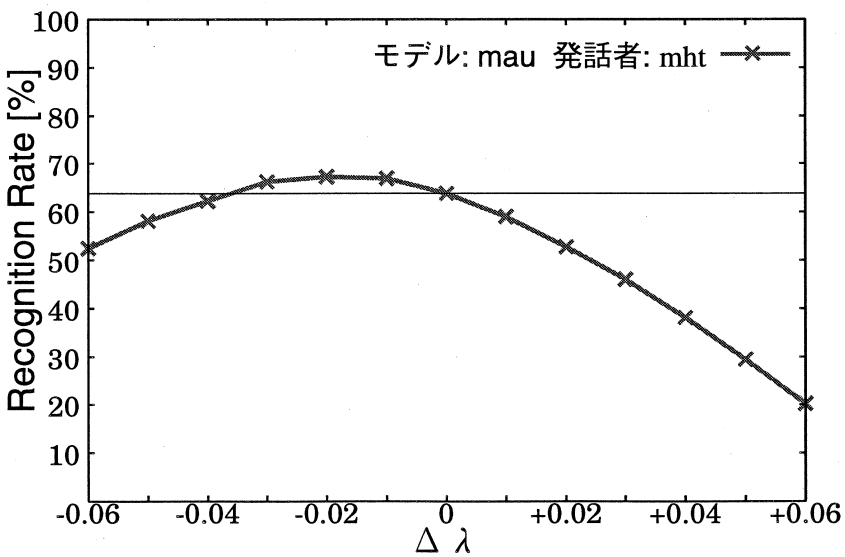

図 3.13: $\Delta\lambda$（話者 mau と話者 mht の変動）と認識率の関係

表 3.2: 適応単語数と雑音 SN 比による $\Delta\lambda$ の推定値

	0dB	10dB	20dB	30dB
8 単語	0.0203	0.0033	−0.0054	−0.0078
16 単語	0.0179	−0.0024	−0.0113	−0.0126
32 単語	−0.0030	−0.0166	−0.0214	−0.0213
64 単語	−0.0265	−0.0257	−0.0259	−0.0229

～4 %，女性話者 (条件 (II)) で 1 %程度である．これは，前述の図 3 の結果から分かるように声道長補正の効果が比較的小さいためであると考えられる．また，図 3.16 より，適応単語数が増えるにつれて認識率が向上する事が確認できた．また，今回は，HMM において平均パラメータのみ適応を行い，分散に関しては適応を行っていない．そこで，平均値のみ再学習したモデルとの比較も行った．図 3.14 より，平均値に関してもまだ適応出来ていない部分があり，さらに分散に関しても適応を行っていく必要があることが分かる．

3.3.5 MLLR との比較実験

参考実験として，話者適応の代表的な手法であり，雑音適応に関しても有効性が知られている MLLR(Maximum Likelihood Linear Regression)[16] との比較実験を行った．MLLR の回帰木のクラスタ数は 16 に設定した．図 3.17 ～ 図 3.20 に各 SNR 毎の比較実験結果を示す．図 3.19, 図 3.20 より，SNR の高いレベル (20,30dB) で少ない単語数 (2～6 単語) では，MLLR よりよい結果が得られたがそれ以外では，MLLR の方が認識率が高いことがわかる．

3.4 考察

本研究では，背景雑音などの加法性歪みや伝達特性などの乗法性歪みに加えて，話者の特性の変化に対しても同時適応が行えるよう，ヤコビ適応法の拡張を行った．話者の特性として周波数伸縮係数 λ を新たに設定し，実験的に正しく推定できることを確認し，認識率を向上した．MLLR との比較では，SNR の高いレベル (20,30dB) で少ない単語数 (2～6 単語) では，MLLR よりよい結果が得られたがそれ以外では，MLLR より低い認識率であった．ヤコビ適応法の特徴としては，少ない処理時間や計算量という特徴が挙げられるが，今後はこの処理時間や計算量についても MLLR との比較を行っていく必要がある．

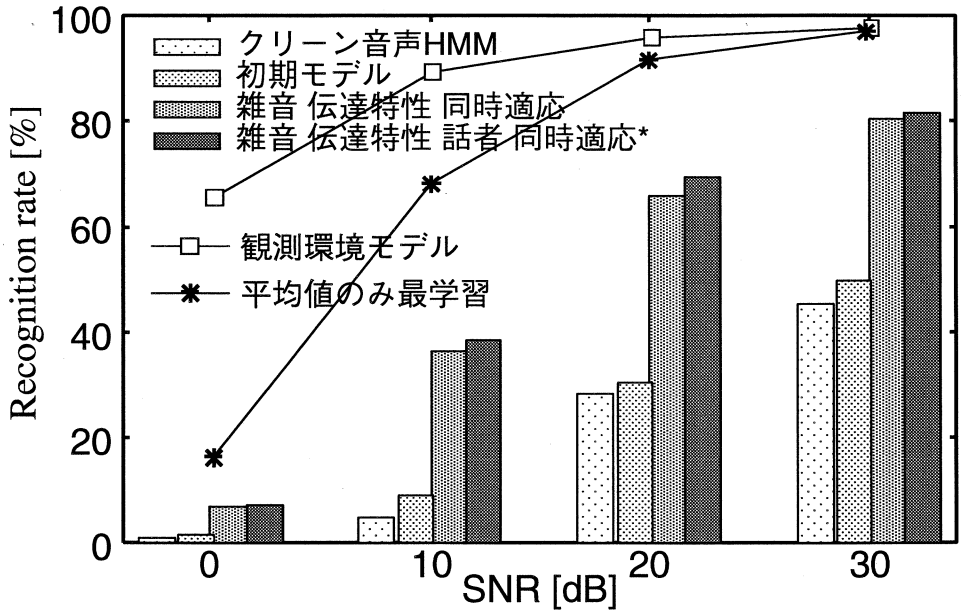

図 3.14: 男性話者における認識結果（条件 (I)，適応単語数 16 単語）

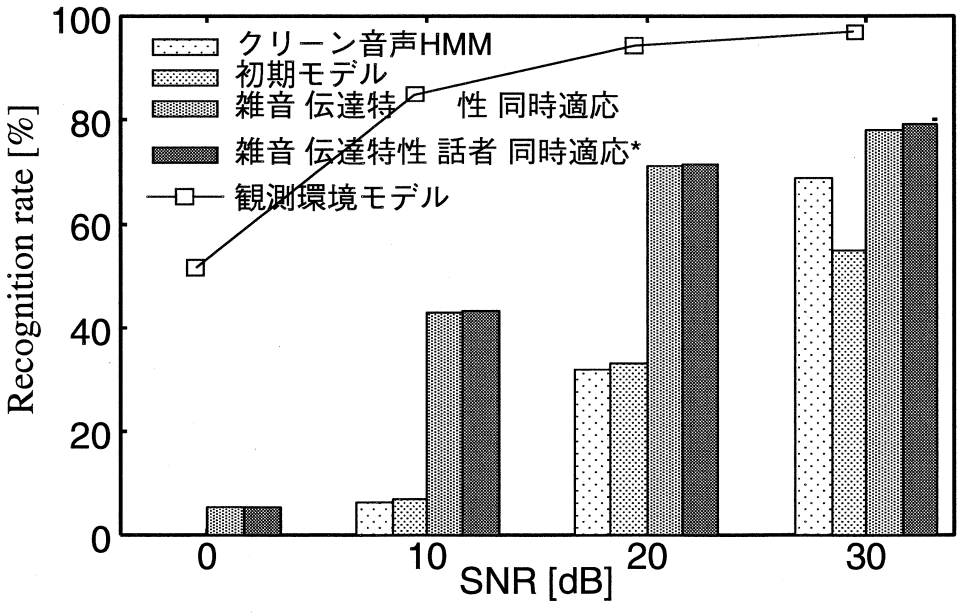

図 3.15: 女性話者における認識結果（条件 (II)，適応単語数 16 単語）

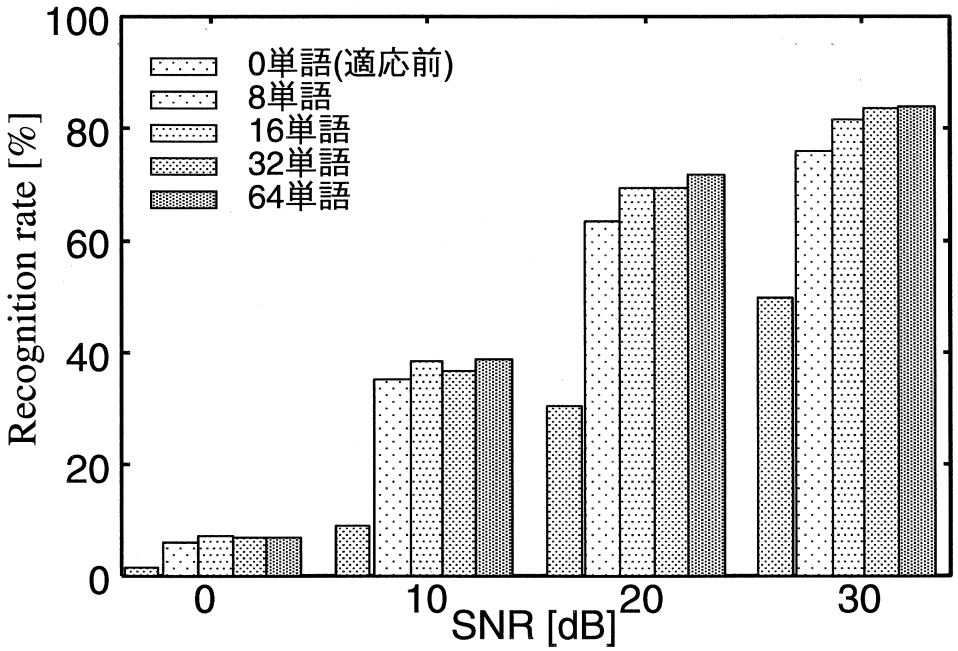

図 3.16: 適応単語数毎の認識結果（条件 (I)）

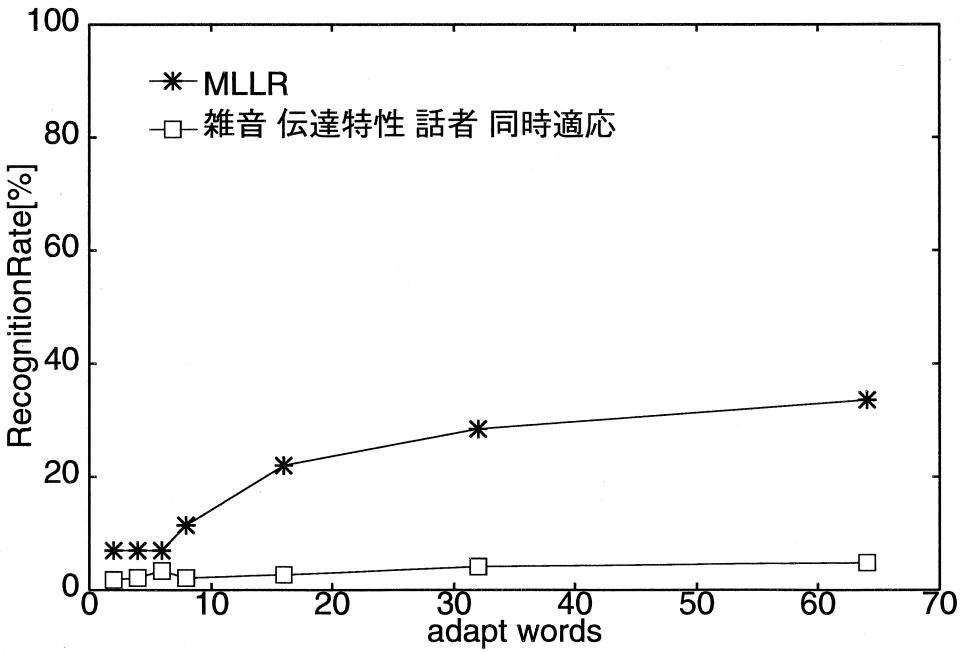

図 3.17: 適応単語数毎の MLLR との比較 (SNR 0dB)

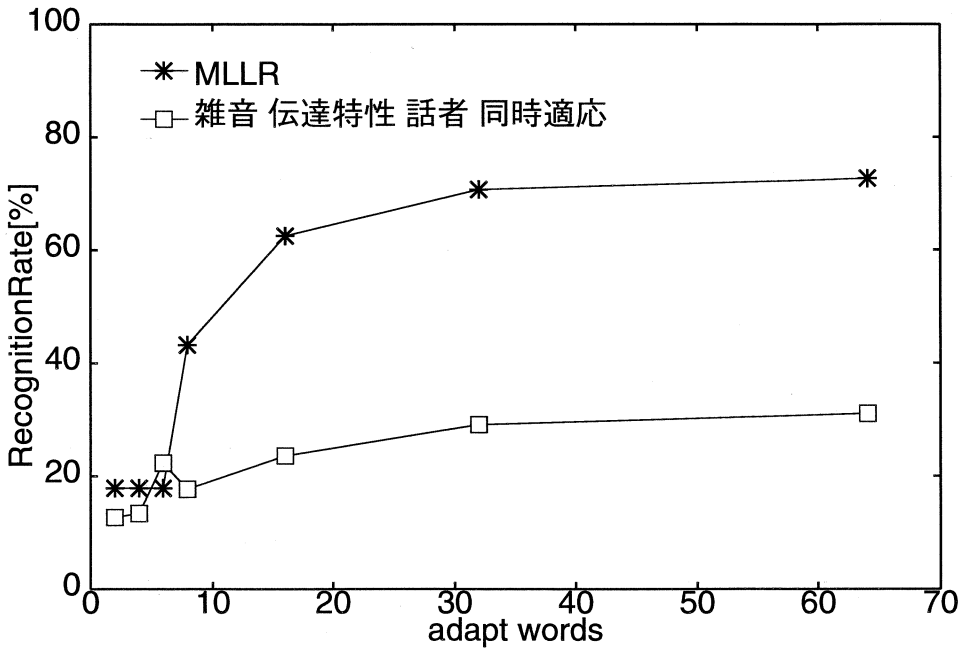

図 3.18: 適応単語数毎の MLLR との比較 (SNR 10dB)

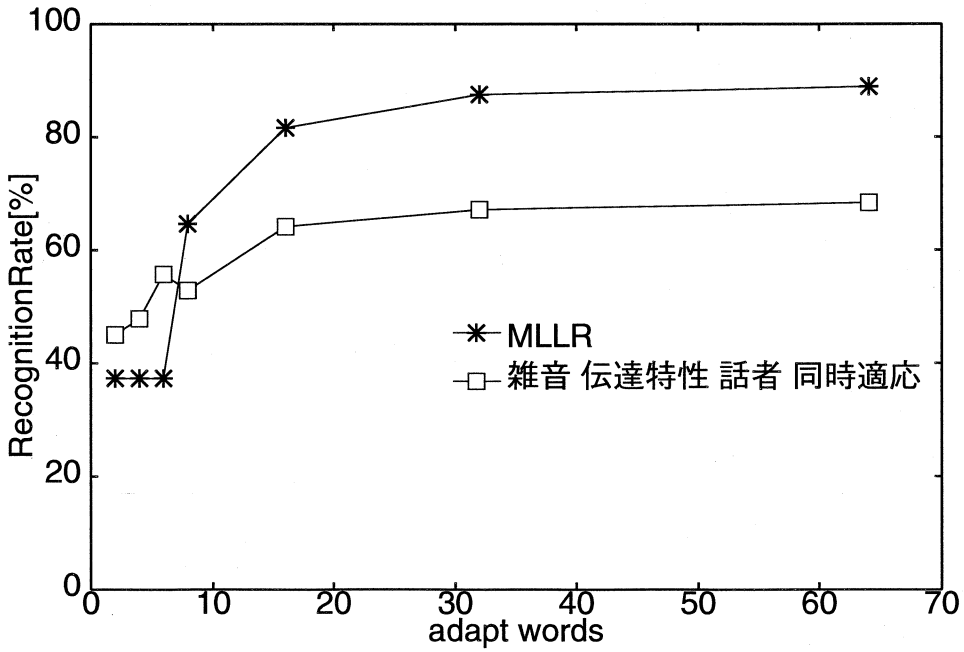

図 3.19: 適応単語数毎の MLLR との比較 (SNR 20dB)

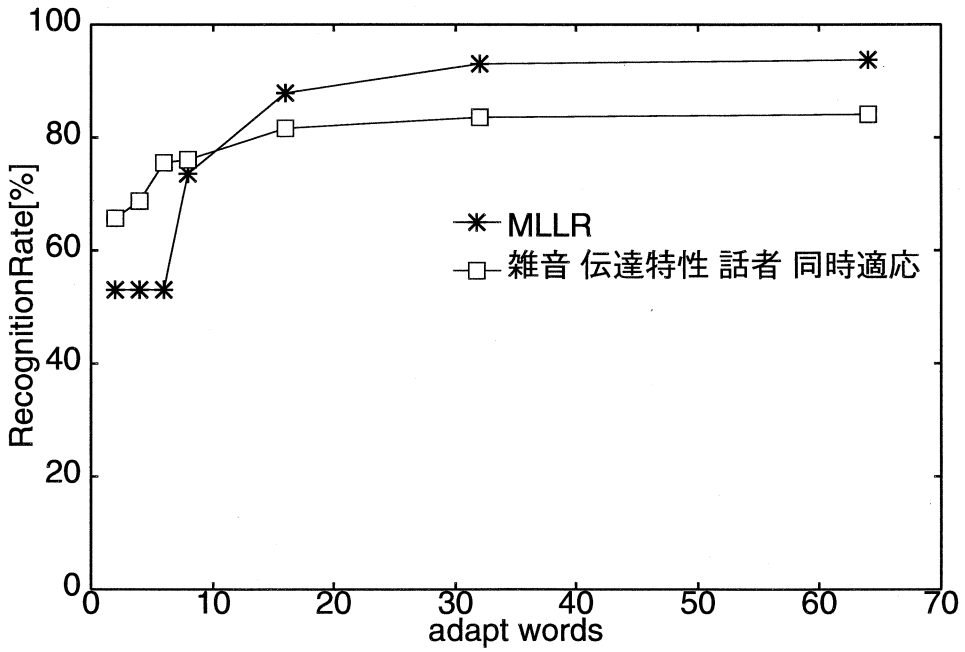

図 3.20: 適応単語数毎の MLLR との比較 (SNR 30dB)

第4章　発声変形に対するヤコビ適応法

高騒音環境下における音声認識では，背景雑音やマイクロフォンの伝達特性に加えて音声そのものの発声変形（Lombard 効果）が問題となり，発声変形に対しても同時に高速に適応する必要がある．本章では，まず発声変形の特徴について調べ，更に話者に対するヤコビ適応法の周波数軸伸縮による手法を改良し，発声変形に対する適応が行えるようにする．

4.1　発声変形に対するヤコビ適応法

4.1.1　発声変形の特徴

背景に高レベル雑音が存在する場合での発声では，雑音に打ち勝つためにより強く発声しようとする．この際の発話した音声の変形は，Lombard 効果として 1911 年に Etienne Lombard によって報告されている [9],[10]．発声時の舌・顎などの調音器官が通常時の発声時と異なる動きを見せるため，Lombard 効果は単に音声のパワーだけでなく複雑な現象となる．

Lombard 効果では音声が主に次のように変化することが報告されている [11].

- ホルマント周波数の移動
- スペクトル傾斜の変化
- ピッチ周波数の移動
- 音韻継続時間の変化

この内，ホルマント周波数の移動は，1.5kHz 付近を境にして，それよりも低い周波数領域にあるものは高域，高い周波数領域にあるものは低域に移動する（図 4.1）．

4.1.2　発声変形モデル

第 4.1.1 節で示した Lombard 効果による変化の内，スペクトルの平均傾斜の変化は，伝達特性の変動として捉えることができ，既に報告した雑音・伝達特性への同時適応での伝

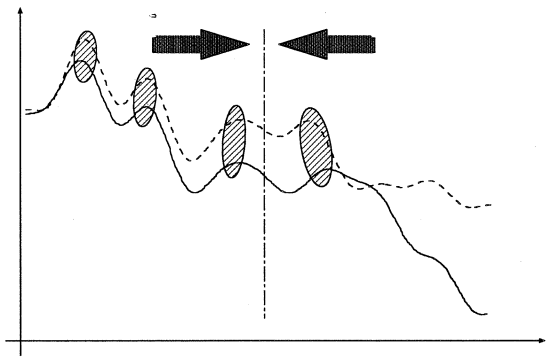

図 4.1: ホルマント周波数の移動の様子

達特性へのヤコビ適応によって対処できる．ここでは，新たにホルマント周波数の移動に対する適応について考える．

第 4.1.1 節でも示したように，Lombard 効果によって，ホルマント周波数は，1.5 kHz 付近を境にして高周波数領域のものは低域へ，低周波数領域のものは高域へ移動すると報告されている [12]．これを 1.5 kHz を中心としたスペクトルの周波数軸伸縮として次のようにモデル化する．通常音声のスペクトル $S(f)$ に対し，周波数軸伸縮後の周波数 f' を周波数軸伸縮係数 λ，周波数定数 f_α を用いて次式で表す．

$$f' = \lambda f + (1 - \lambda) f_\alpha \tag{4.1}$$

ここで，$\lambda > 1$ は周波数 f_α に向かう収縮を意味し，$\lambda < 1$ の場合は逆に伸長する．ここでのスペクトルの変化と周波数 f' の線形変換の様子は図 4.2 のようになる．

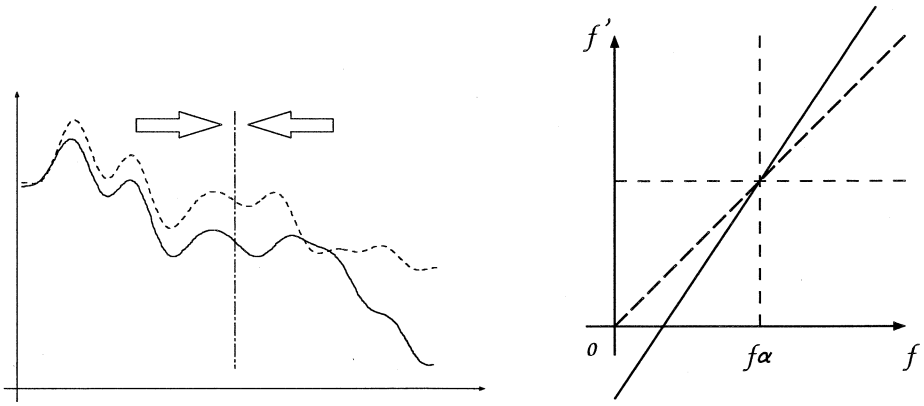

図 4.2: スペクトルの変化と周波数 f' の線形変換の様子

4.1.3 発声変形モデルを用いたヤコビ適応

音声スペクトル $S(f)$ のケプストラムを $\boldsymbol{C}$，$S(f')$ のケプストラムを $\boldsymbol{C}_L$ とすると，それらは次式の関係にある．

$$\boldsymbol{C}_L = \boldsymbol{F}^{-1}\boldsymbol{F}_L\boldsymbol{C} \tag{4.2}$$

ここで，$\boldsymbol{F}^{-1}$ は離散コサイン変換行列（$\boldsymbol{F}$）の逆行列，$\boldsymbol{F}_L$ は前述の周波数軸伸縮を伴う離散コサイン変換行列で，その (i,j) 成分は次式で与えられる．

$$(\boldsymbol{F}_L)_{ij} = \cos\frac{i(\lambda(j+0.5)+(1-\lambda)\omega_\alpha)\pi}{N} \tag{4.3}$$

N は周波数の標本化点数，ω_α は f_α に対応する周波数軸定数である．

発声変形によって，伸縮係数 λ が微少量 $\Delta\lambda$ だけ変化したとすると，ケプストラム $\boldsymbol{C}_L$ の変動量 $\Delta\boldsymbol{C}_L$ は，Taylor 展開による線形近似によって次式で与えられる．

$$\Delta\boldsymbol{C}_L = \frac{\partial\boldsymbol{C}_L}{\partial\lambda}\Delta\lambda \tag{4.4}$$

ここで，$\boldsymbol{J}_L \equiv \frac{\partial\boldsymbol{C}_L}{\partial\lambda}$ と置き，これを発声変形ヤコビ行列と呼ぶ．$\lambda \approx 1$ のとき，$\boldsymbol{J}_L$ は (4.2) 式より，

$$\boldsymbol{J}_L = \boldsymbol{F}^{-1}\frac{\partial\boldsymbol{F}_L}{\partial\lambda}\boldsymbol{C} \approx \boldsymbol{F}^{-1}\boldsymbol{G}\boldsymbol{C} \tag{4.5}$$

となり，$\boldsymbol{G}$の (i,j) 成分は

$$(\boldsymbol{G})_{ij} = \frac{-i(j+0.5-\omega_\alpha)\pi}{N}\sin\left(\frac{i(j+0.5)\pi}{N}\right) \tag{4.6}$$

である．この $\boldsymbol{J}_L$ は想定する初期環境で事前に求まる．

4.2 Lombard 音声収録

4.2.1 予備実験

収録を始めるにあたって，どのようなデータをどれだけ取れば良いのかを把握する必要がある．まず，背景雑音としては，電子協の騒音データベースから，駅構内雑音 (No.05)

学習単語数	認識率
ATRAset 655words	98.32
ATRAset 393words	97.86
ATRAset 262words	エラー (音素/p/データ足りず)
ATRAset balance216words	90.69

表 4.1: 学習単語数による認識率比較

通常音声	mnsc	mmsc	mhmc	mhnc
Lombard 音声	mnsl	mmsl	mhml	mhnl

表 4.2: 音声データ ID

を用いた．これは，比較的低域にエネルギーを持つため，Lombard 効果が起こり易い為である．

また，収録単語は最低限どのくらい必要かを調べるために，学習単語の数を表 4.1 のように変化させて，認識率を比較してみた．

この予備実験から，655 単語，393 単語では，高い認識率を示しているが，262 単語では，音素 /p/ の学習データが足りず，認識率が出せなかった．バランス単語 216 単語では，認識率は出たが，90.69 % と低い値となった．

そこで，学習用として 393 単語，認識用として 262 単語，計 655 単語を通常音声と Lombard 音声それぞれ収録することにした．

4.2.2 Lombard 音声の収録方法

収録は成人男性 4 名，それぞれ通常音声と Lombard 音声とを収録した．通常音声は，約 30dBSPL の室内で収録し，Lombard 音声は，発話者がイヤホンから高レベル雑音 (約 60dBSPL) を聞きながら，マイクに向かって発話した音声を収録した．通常音声と Lombard 音声それぞれに対して，学習用として 393 単語，認識用として 262 単語の計 655 単語を収録した．この際，発声した音声のイヤホンへのフィードバックは無い．雑音は電子協騒音データベースから駅構内雑音 (No.05) を用いた．録音機材としては，エレクトレットコンデンサーマイク (SONYECM-S959C)，DAT 録音機 (SONY TCD-D100) を用いた．4 名の通常音声データ，Lombard 音声データに対して表 4.2 のような ID を付け，以降この ID を用いることにする．最初の 1 文字の m は男性 (male) を表し，最後の 1 文字は，通常音声 (clean) か Lombard 音声 (lombard) かを表す．

表 4.3: 実験条件

発声者	成人男性 4 名
通常音声	室内（約 30 dBSPL）
Lombard 音声	駅構内（約 60 dBSPL）
	（電子協騒音データを聴きながらの擬似環境）
収録単語	655 単語
学習資料	393 単語
評価資料	262 単語 （学習資料以外）
辞書内単語	655 単語
特徴量	16 次 LPC + 16 次 Δ LPC （0 次を含む）
音素 HMM	音素環境独立型， 3 状態， 3 混合

表 4.4: 各話者毎の ω_α による認識率

ω_α	2	4	6	8	9	10	11	12	13	14	16
mnsl	57.63	59.16	63.36	67.56	69.08	70.99	71.37	**71.76**	**71.76**	70.99	70.99
mmsl	49.24	50.38	52.29	56.11	56.49	**56.87**	56.49	56.49	56.11	54.58	52.29
mhml	67.94	69.08	70.99	**72.90**	**72.90**	72.14	70.99	69.85	69.08	69.47	69.08
mhnl	81.30	**81.68**	80.53	81.30	**81.68**	81.30	80.92	80.92	81.30	80.53	80.15

4.3 認識実験

提案手法の性能を調べるために, 収録した Lombard 音声を用いて，孤立単語音声認識実験を行った．実験条件を，表 4.3 に示す．

4.3.1 周波数軸定数 ω_α に関する実験

4.1.2 で提案した発声変形モデルでは，ホルマント周波数は，ω_α を境にして，それよりも低い周波数領域にあるものは高域，高い周波数領域にあるものは低域に移動する．この ω_α の最適値を調べるために，ω_α の値を表 4.4 のように変化させて，認識率を比較した．ここで，$\Delta\lambda$ の値は最小二乗法を用いて求めた (次節参照)．適応単語数は 16 単語である．

表 4.4 より，話者 mnsl では $\omega_\alpha = 12, 13$ ，話者 mmsl では $\omega_\alpha = 10$ ，話者 mhml では $\omega_\alpha = 8, 9$ ，話者 mhnl では $\omega_\alpha = 4, 9$ ，が ω_α の最適値であることがわかる．最適な ω_α は，話者によってある程度ばらつきがあることが確認できた．

4.3.2 発声変形モデルによるヤコビ適応実験

発声変形モデルによるヤコビ適応の認識実験を行い，その効果を調べた．発声変形モデルによるヤコビ適応では，$\Delta\lambda$の値が必要になる．そこで，いくつかの発声サンプルから，最小二乗法により$\Delta\lambda$の値を推定する．詳細なアルゴリズムや処理の流れは，5章で説明する．図4.3が適応単語数16単語の場合の認識結果を示す．図の上にあるcloseの線は観測したLombard音声そのものによって学習したモデルによる認識結果である．図4.3より，ω_αの値を最適値にした場合，2%弱の認識率改善が見られた．

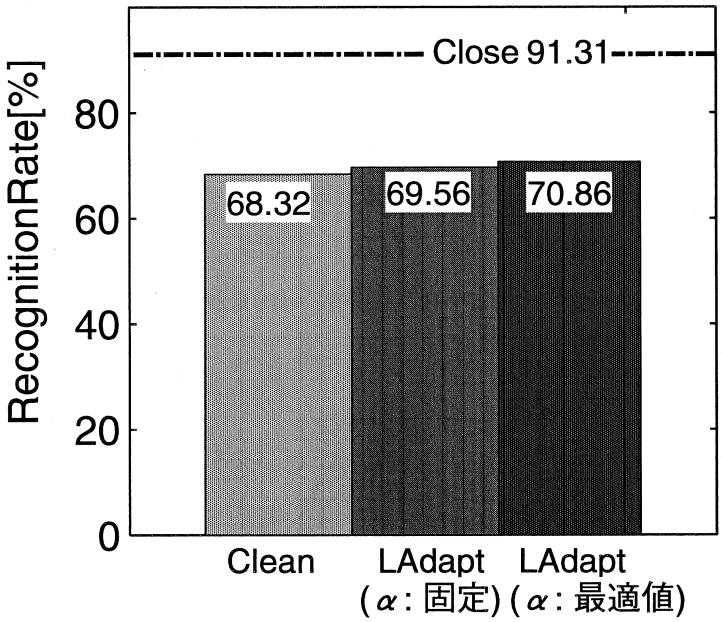

図 4.3: 単語認識率の比較（適応単語数 16）

4.3.3 加法性・乗法性も考慮した同時適応実験

発声変形においてもスペクトルの加法性・乗法性の適応の効果があるかどうかを調べるために，加法性・乗法性に対する同時適応を雑音・伝達特性同時適応を用いて行った．さらに，発声変形モデルも加えた加法性・乗法性・発声変形モデルに対する同時適応を雑音・伝達特性・発声変形同時適応 (5章参照) を用いて行った．(ただし，雑音を重畳していないため，仮定した初期環境における雑音ヤコビ行列$\boldsymbol{J}_N$を計算できないので，ここでは前節の実験で使用した雑音ヤコビ行列を代用した．)

図4.4の認識結果より，雑音・伝達特性の2つの同時適応だけでもかなりの認識率の向上が見られることから，Lombard効果には加法性・乗法性とみなせる変形要素が大きいと考えられる．

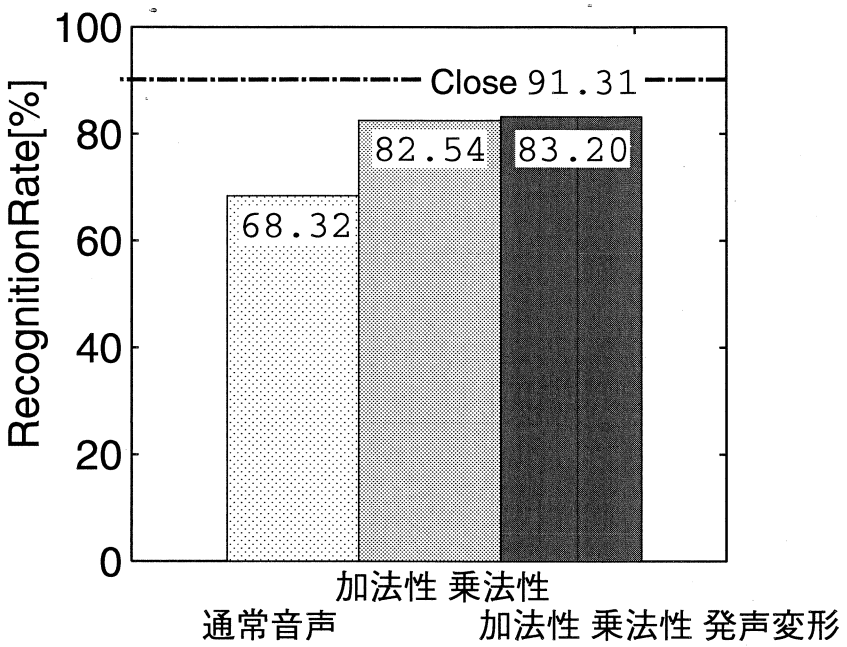

図 4.4: 単語認識率の比較（適応単語数 16）

4.4 考察

周波数軸定数ω_αの最適値を調べるために，ω_αの値を変化させて認識率の変化を調べた．周波数軸定数ω_αの最適値は，話者によってある程度ばらつきがあることがわかた．ω_αに関してもλと同時に推定する手法を検討する必要がある．また，発声変形モデルを用いたヤコビ適応では，最適なω_αを用いた場合でも，認識率の改善は僅かであったが，雑音・伝達特性の 2 つの同時適応では，かなりの認識率の向上が見られたことから，Lombard 効果には加法性・乗法性とみなせる変形要素が大きいと考えられる．

第5章　ヤコビ適応法を用いた雑音・伝達特性・発声変形への同時適応

高騒音環境下における音声認識では，背景雑音やマイクロフォンの伝達特性に加えて音声そのものの発声変形（Lombard 効果）が問題となり，発声変形に対しても同時に高速に適応を行う必要がある．

4 章では，発声変形に対するヤコビ適応法を提案し，(雑音・伝達特性を付加していない)Lombard 音声に対して適応を行って，実験によりその効果を調べた．本章では，雑音・伝達特性を付加した Lombard 音声に対して，雑音・伝達特性・発声変形への同時適応を行い，実験を通してその効果を確認する．

5.1　雑音・伝達特性・発声変形への同時適応

5.1.1　基本原理

3.2 節では，雑音，伝達特性，話者に対する同時適応を最小二乗法を用いて行った．本節では，同様に最小二乗法を用いて，雑音，伝達特性，発声変形の 3 つの変動に対して同時適応を行う．ここでは図 5.1 に示す観測系を想定する．仮定した初期モデルから発話環境モデルへのケプストラムの変動量 $\Delta \boldsymbol{C}_Y$ は，発声変形ヤコビ行列 $\boldsymbol{J}_L$，雑音ヤコビ行列 $\boldsymbol{J}_N$ を用いて，次式で表せる．

$$\Delta \boldsymbol{C}_Y = \boldsymbol{J}_L \Delta\lambda + \boldsymbol{J}_N \Delta \boldsymbol{C}_N + \Delta \boldsymbol{C}_H \tag{5.1}$$

$$\boldsymbol{J}_N \equiv \frac{\partial \boldsymbol{C}_Y}{\partial \boldsymbol{C}_N} = \boldsymbol{F}^{-1} \frac{\boldsymbol{S}_N}{\boldsymbol{S}_S + \boldsymbol{S}_N} \boldsymbol{F} \tag{5.2}$$

$\Delta\lambda$，$\Delta \boldsymbol{C}_N$，$\Delta \boldsymbol{C}_H$ はそれぞれ周波数軸伸縮係数の変動量，雑音の変動量，伝達特性の変動量であり，観測環境下で発声された音素列の既知な数単語を用いて推定できる．

音素 HMM のパラメータである i 番目の正規分布の平均ベクトルについて，初期環境と観測環境の変動量を $\Delta \overline{\boldsymbol{C}_Y}^{(i)}$，観測に伴う誤差ベクトルを $\boldsymbol{\epsilon}^{(i)}$ とする．全ての音素モデルの総分布数を M とすると，推定すべき $\Delta\widehat{\lambda}$，$\Delta\widehat{\boldsymbol{C}_N}$，$\Delta\widehat{\boldsymbol{C}_H}$ は以下の M 個の連立方程式の誤差ベクトルの自乗和を最小化することによって得られる．

$$
\begin{cases}
\Delta\overline{C_Y}^{(1)} = \boldsymbol{J}_L^{(1)}\Delta\hat{\lambda} + \boldsymbol{J}_N^{(1)}\Delta\widehat{C_N} + \Delta\widehat{C_H} + \boldsymbol{\epsilon}^{(1)} \\
\Delta\overline{C_Y}^{(2)} = \boldsymbol{J}_L^{(2)}\Delta\hat{\lambda} + \boldsymbol{J}_N^{(2)}\Delta\widehat{C_N} + \Delta\widehat{C_H} + \boldsymbol{\epsilon}^{(2)} \\
\quad\vdots \\
\Delta\overline{C_Y}^{(M)} = \boldsymbol{J}_L^{(M)}\Delta\hat{\lambda} + \boldsymbol{J}_N^{(M)}\Delta\widehat{C_N} + \Delta\widehat{C_H} + \boldsymbol{\epsilon}^{(M)}
\end{cases}
\tag{5.3}
$$

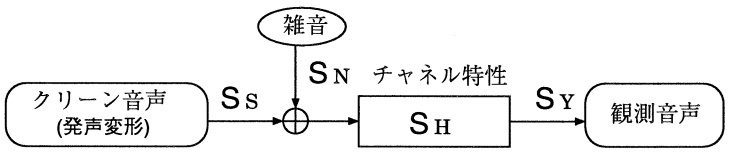

図 5.1: 仮定する音声信号の観測系

5.1.2 適応アルゴリズム

実際の適応処理の流れは図5.2に示される通りで，その処理は以下の事前処理・適応処理・認識処理の3段階に分けられる．

事前処理

(1) $\boldsymbol{J}_L$ の計算： 周波数定数 f_α を決定し，発声変形ヤコビ行列を計算．

(2) 初期環境モデルの作成： 仮定した雑音環境 A に基づき，HMM 合成法により雑音重畳音声 HMM ($\overline{\boldsymbol{C}_Y}^A$) を作成．

(3) $\boldsymbol{J}_N$ の計算： 仮定した雑音と雑音重畳音声から雑音ヤコビ行列を計算．

適応処理

(1) 観測環境モデルの作成： 観測環境 B で収集した音声資料を初期モデルで Viterbi セグメンテーションする事により，少数個の音素 HMM ($\overline{\boldsymbol{C}_Y}^B$) を作成．

(2) 変動量の推定： 観測された音素について初期環境モデル A と観測環境モデル B の平均ベクトルの変動量 ($\Delta\overline{C_Y} = \overline{\boldsymbol{C}_Y}^B - \overline{\boldsymbol{C}_Y}^A$) を計算し，最小自乗法を用いて $\Delta\widehat{C_N}$, $\Delta\widehat{C_H}$, $\Delta\hat{\lambda}$ を推定．

(3) モデルの適応： 初期環境モデル A のパラメータを修正する事により，全ての音素 HMM を適応．

認識処理

(1) 認識：適応したHMMモデル を用いて，観測環境 の音声に対する認識を行う

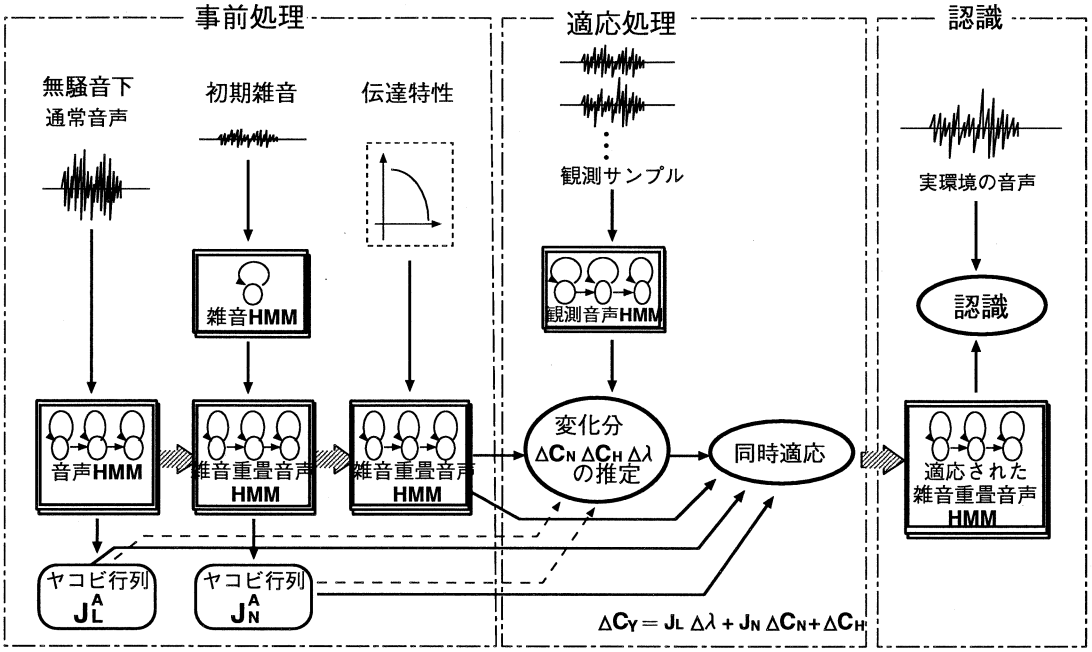

図 5.2: 雑音・伝達特性・発声変形へのヤコビ同時適応システム処理の流れ

5.2 評価実験

共通する実験条件を表5.1に示す．音声資料として，室内で発声した通常音声と，イヤホンにより高レベルの雑音（電子協・駅構内）を聞きながら発声したLombard音声を雑音と分離して収集した．

5.2.1 適応の様子

実際にHMM合成法(PMC法)によって作成した初期モデルから，雑音・伝達特性・発声変形への同時適応により，どのようにスペクトルが変化するか，その変化の様子を図5.3～5.10に示す．図中のPMCは初期モデルのスペクトルを表し，adapt(8words),adapt(8words)は，初期モデルから適応単語数8words,64wordsにおける同時適応後のスペクトルを表し，closeは観測環境の音声によって学習した理想モデルのスペクトルになる．図5.3～5.10により，適応単語数を増やすにつれて，初期モデルから，観測環境モデルに近づいている様子が確認できた．

表 5.1: 実験条件

発声者	成人男性 4 名
通常音声	室内（約 30 dBSPL）
Lombard 音声	駅構内（約 60 dBSPL）
	（電子協騒音データを聴きながらの擬似環境）
収録単語	655 単語
学習資料	393 単語
評価資料	262 単語 （学習資料以外）
辞書内単語	655 単語
特徴量	16 次 LPC + 16 次 Δ LPC （0 次を含む）
音素 HMM	音素環境独立型， 3 状態， 3 混合

表 5.2: 仮定する初期環境と認識を行う観測環境
条件 1 : 雑音・伝達特性を付加した Lombard 音声に対する適応実験

	初期環境		観測環境
発声	通常音声	⇒	Lombard 音声
雑音重畳	駅構内		駅構内
伝達特性	全域通過	⇒	低域通過 (図 3.3)

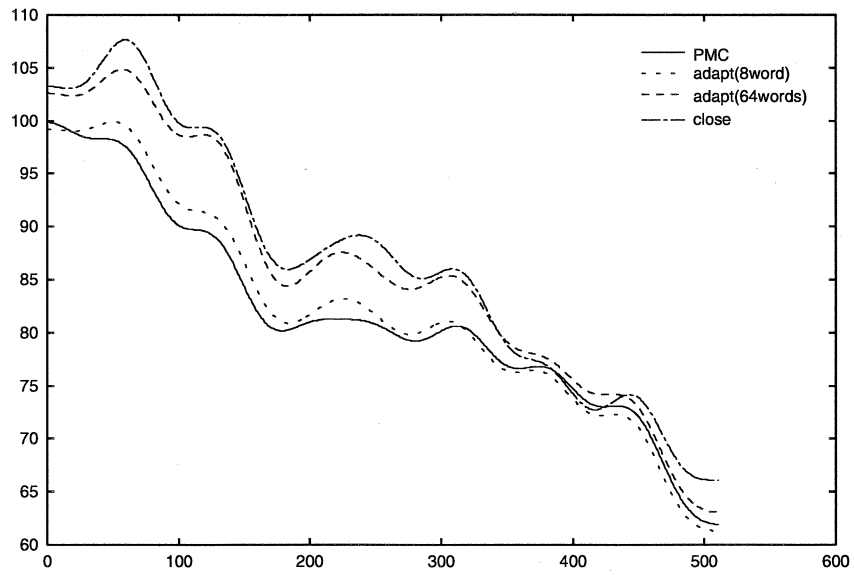

図 5.3: 音素 /a/ 話者 mns CNL 同時適応の様子

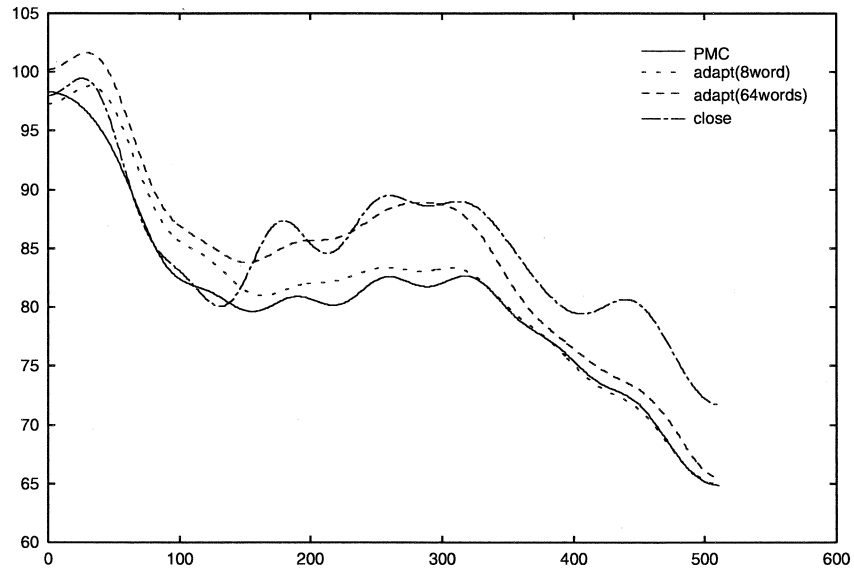

図 5.4: 音素 /i/ 話者 mns CNL 同時適応の様子

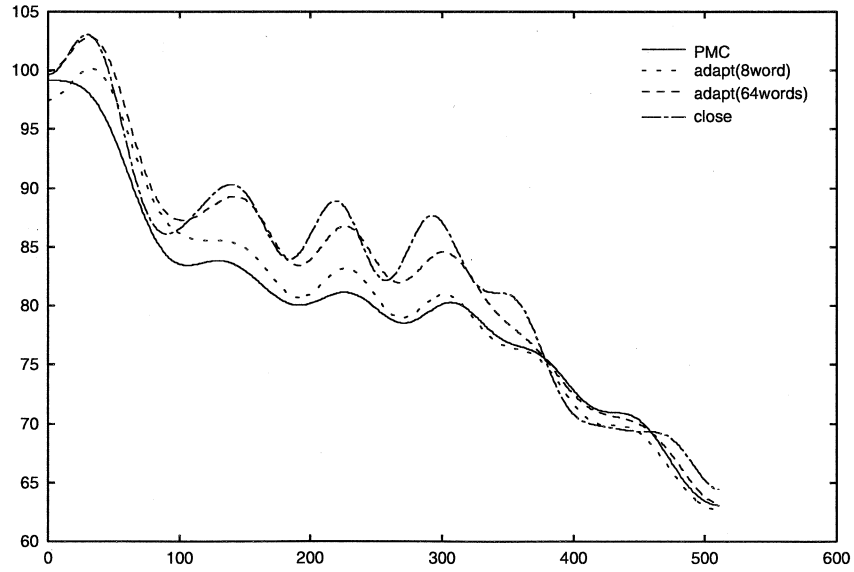

図 5.5: 音素 /u/ 話者 mns CNL 同時適応の様子

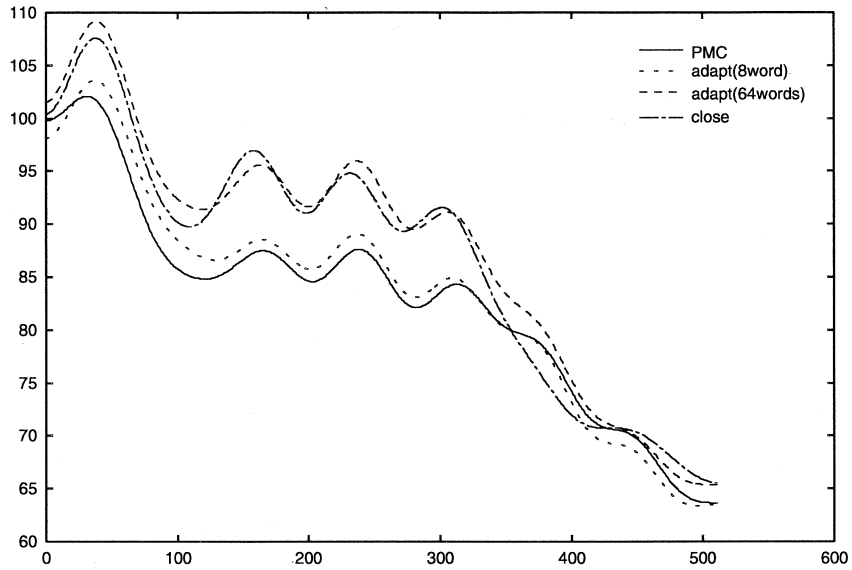

図 5.6: 音素 /e/ 話者 mns CNL 同時適応の様子

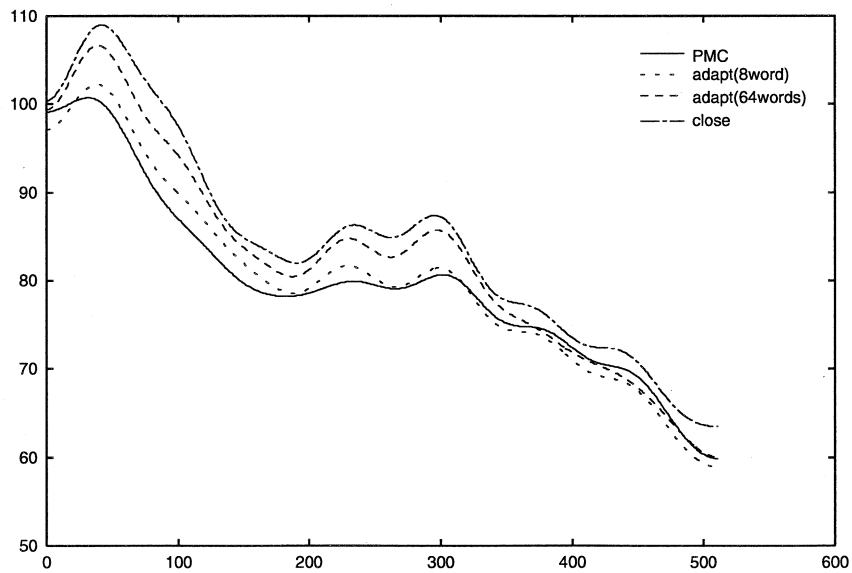

図 5.7: 音素 /o/ 話者 mns CNL 同時適応の様子

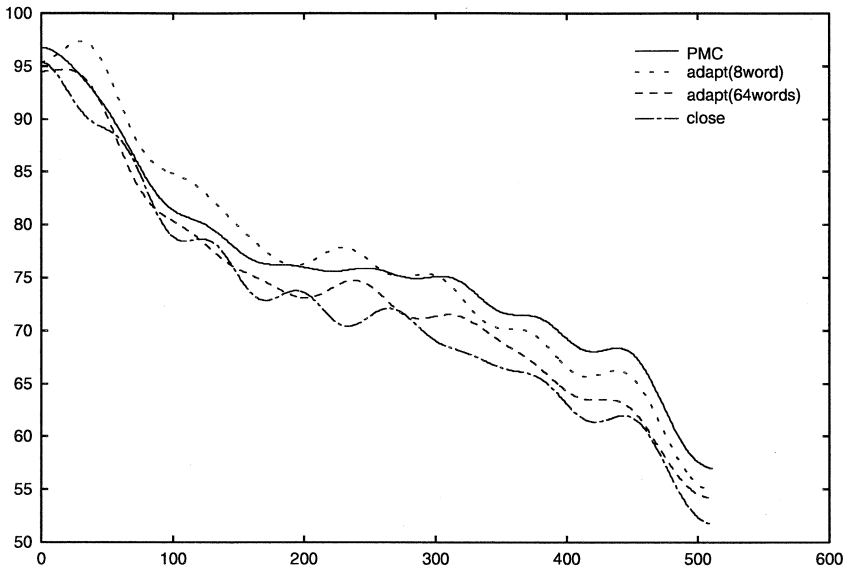

図 5.8: 音素 /p/ 話者 mns CNL 同時適応の様子

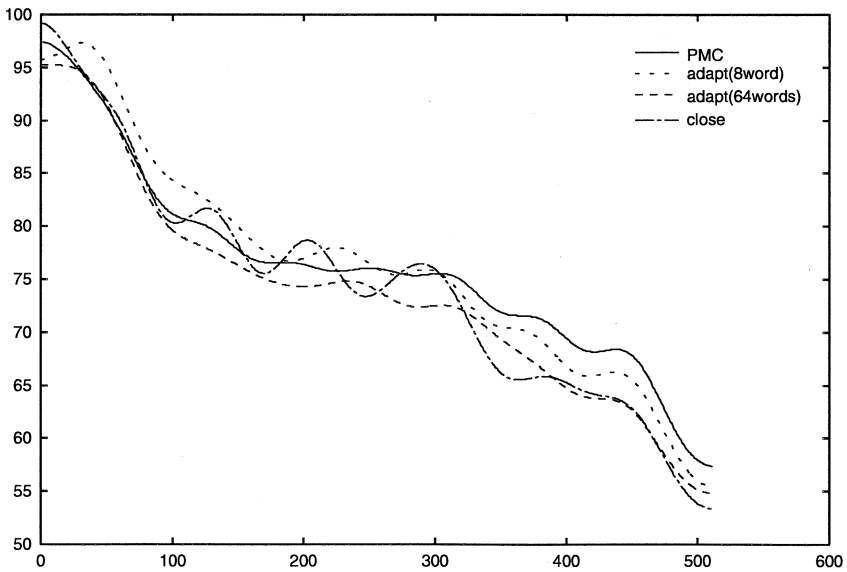

図 5.9: 音素 /b/ 話者 mns CNL 同時適応の様子

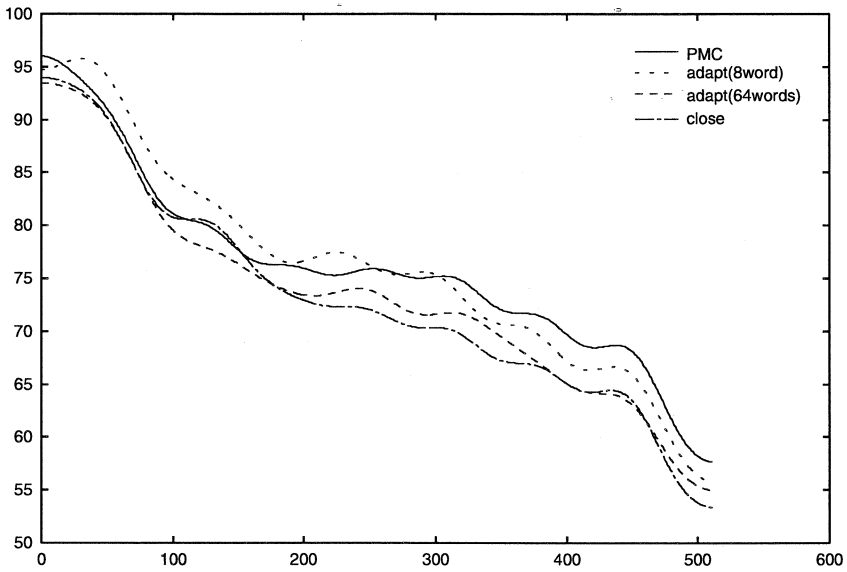

図 5.10: 音素 /k/ 話者 mns CNL 同時適応の様子

5.2.2 発声変形・雑音・伝達特性への同時適応実験

表5.2の条件の下で同時適応実験を行った．雑音重畳モデルは通常音声で学習した音素HMM と駅構内雑音で学習した 1 状態 1 分布の雑音 HMM を用いて HMM 合成法により作成した．周波数定数 f_α は1話者による予備実験により決定し，全話者共通の値として使用した．

図5.11，5.12 は話者 4 名の平均単語認識率である．図5.11から，適応要因数を増やすにつれて認識率が向上しており，周波数軸伸縮モデルの効果がある事が確認できた．また，図5.12より，適応単語数の増加によって，即ち観測される音素の種類が増えるにつれて，認識率が向上する事も確認できた．

5.3 考察

発声変形におけるホルマント周波数の移動を周波数軸の伸縮としてモデル化し，高騒音環境下での雑音・伝達特性との同時適応法を提案した．適応要因数を増やすにつれて認識率が向上し，周波数軸伸縮モデルの効果を確認できた．また，適応単語数の増加によって，即ち観測される音素の種類が増えるにつれて，認識率が向上する事も確認できた．

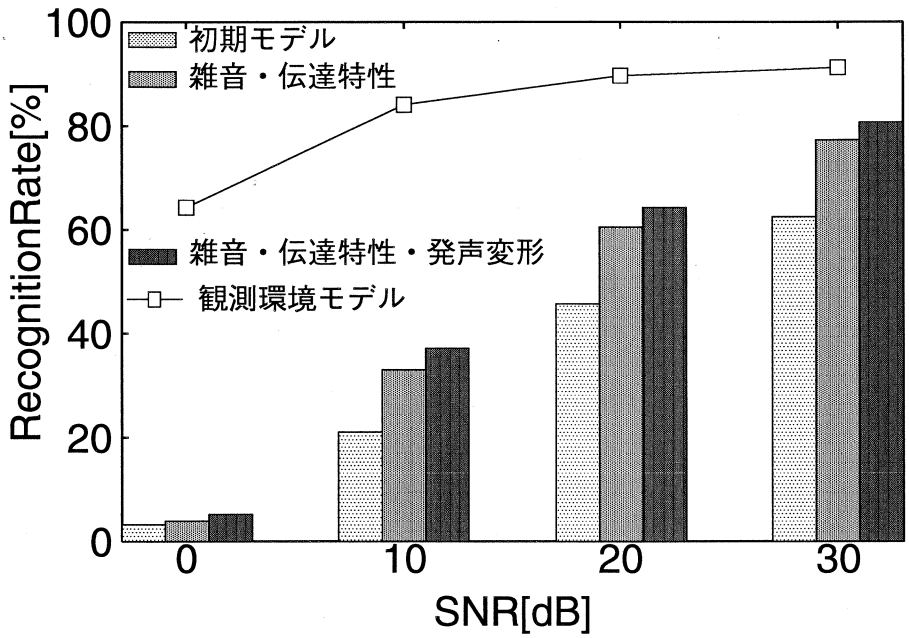

図 5.11: 条件 1 の単語認識率の比較（適応単語数 16）

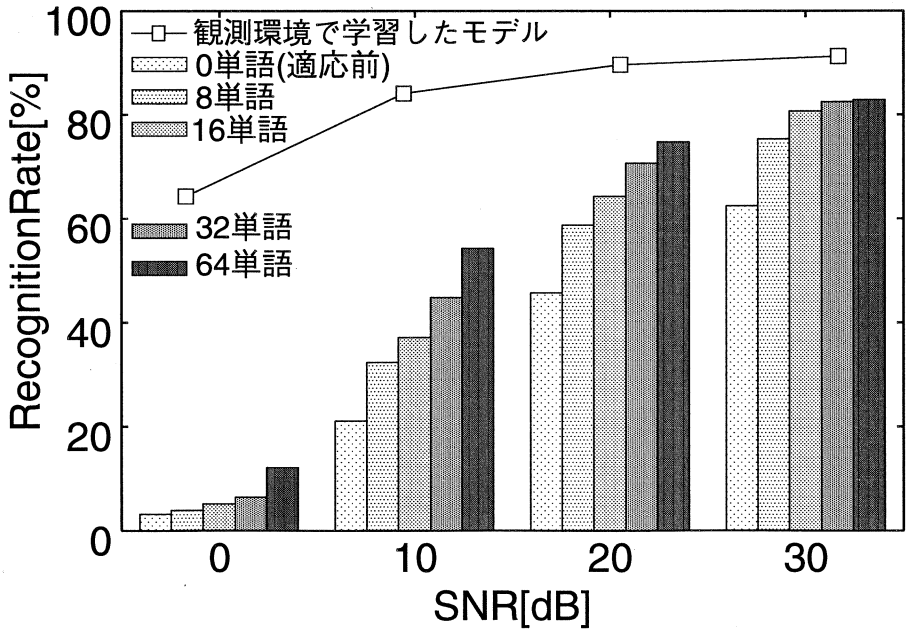

図 5.12: 条件 1 の適応単語数による単語認識率の比較

第6章　結論と今後の課題

6.1　結論

本研究では，まず，背景雑音などの加法性歪みや伝達特性などの乗法性歪みに加えて，話者の特性の変化に対しても同時適応が行えるよう，ヤコビ適応法の拡張を行った．話者の特性として周波数伸縮係数 λ を新たに設定し，実験的に正しく推定できることを確認し，認識率を向上した．さらにその手法を拡張し，発声変形におけるホルマント周波数の移動をある周波数軸定数 ω_α に対しての周波数軸の収縮としてモデル化し，高騒音環境下での雑音・伝達特性との同時適応法を提案した．また，Lombard 効果による発声変形では，周波数軸の収縮以外にも加法性・乗法性の変形が考えられ，雑音・伝達特性の同時適応による認識率の改善を確認した．

6.2　今後の課題

今後の課題としては，雑音・伝達特性・話者同時適応に関しては処理時間・計算量について MLLR との比較を行う．また，発声変形に対する適応に関しては，本研究ではモデルにおける周波数軸定数 ω_α は，予め固定した値を与えているが，その最適値を求める手法について検討する．また，本研究では，話者適応と発声変形適応を別々に行ったがこれらを同時に組み込んだ雑音・伝達特性・話者・発声変形に対する同時適応についても検討する．

謝辞

本研究に対して熱心な御指導，御教示を賜りました，北陸先端科学技術大学院大学 情報科学研究科 下平 博助教授に深く心から感謝致します．

本研究を進めていく上で，多大なる御指導，御鞭撻を賜りました，同研究科 嵯峨山 茂樹教授に深く心から感謝致します．

また，研究内容から本研究の使用ツール類などに関して，幅広く御助言を賜りました，同研究科 中井 満助手に深く心から感謝致します．

また，日頃から本研究に対する貴重な御意見，御助力を賜りました，同研究科 博士後期課程 六井 淳氏，松田 繁樹氏，藤永 勝久氏に深く心から感謝致します．

最後に，本研究において必要となった実験に快く御協力していただき，また有意義な研究生活を送るために心の支えとなってくれた，嵯峨山・下平研究室の皆様方に深く心から感謝致します．

関連図書

[1] J. C. Junqua, "The Lombard reflex and its role on human listeners and automatic speech recognizer," J. Acoustic. Soc. Amer., Vol.93, pp.510-524, 1993.

[2] 山口 義和, 高橋 淳一, 高橋 敏, 嵯峨山 茂樹, "Taylor 展開に基づく高速な音響モデル適応法," 日本音響学会講演論文集, 2-Q-11,pp. 151-152, Sep. 1996.

[3] 山口 義和, 高橋 淳一, 高橋 敏, 嵯峨山 茂樹, "Taylor 展開による音響モデルの適応," 電子情報通信学会 技術研究報告, SP96-78, pp. 1-8, Dec. 1996.

[4] S. Sagayama, Y. Yamaguchi, S. Takahashi, "Jacobian Adaptation of Noisy Speech Models," Proc. ASRU97, pp. 396-403, 1997.

[5] 嵯峨山 茂樹, 山口 義和, 高橋 敏," Jacobi 行列を用いた音響モデルの適応アルゴリズム," 日本音響学会講演論文, 1-6-13, pp. 31-32, Mar. 1997.

[6] 山口 義和, 高橋 敏, 嵯峨山 茂樹, "Jacobian 適応法による雑音適応の性能評価," 日本音響学会講演論文, 1-6-14, pp. 33-34, Mar. 1997.

[7] 赤江俊彦, 中井満, 下平博, 嵯峨山茂樹, "雑音環境へのヤコビ適応法の拡張", 音響論集, 1-8-4, pp.7-8, 2000-3.

[8] 加藤， 他，"ヤコビ適応法を用いた雑音環境と伝達特性への同時適応" 平 12 秋音講論， 1-5-9， pp.17–18 (2000-9).

[9] E. Lombard, "Le signe de l'elevation dela voix," Ann. Maladiers Oreille, Larynx, Nez, Pharynx, vol.37, pp.101-119, 1911.

[10] H.L. Lane and B. Tranel, "The Lombard sign and the role on hearing in speech," Journal of Speech and Hearing Research, vol.14, pp.677-709, 1971.

[11] D. B. Pisoni, et al., "Some Acoustic-Phonetic Correlates of Speech Produced in Noise," Proc. ICASSP, pp.745–748 (1986).

[12] 滝沢, 他, "雑音下での発声変形を考慮した認識方式の検討," 平元秋音講論, 2-1-5, pp.61–62 (1989-10).

[13] Lawrence Rabiner, Biing-Hwang Juang, "音声認識の基礎," NTT アドバンスドテクノロジ株式会社, 1995.

[14] E. Eide and H. Gish, "A Parametric Approach to Vocal Tract Length Normalization," Proc. ICASSP-96, Vol. 1, pp. 346-349, 1996.

[15] J. L. Gauvain and C. H. Lee, "Maximum a Posteriori Estimation for Multivariate Gaussian Mixture Oservations of Markov Chains," IEEE Trans. Speech Audio Process., Vol.2, pp. 291-298, 1994.

[16] M. J. F. Gales and P. C. Woodland, "Mean and Variance Adaptation within the MLLR framework," Computer Speech and Language, Vol.10, pp.249-264, 1996.

[17] M. J. F. Gales and S. J. Young, "An Improved Approach to the Hidden Markov Model Decomposition of Speech and Noise," Proc. ICASSP92, pp.2 33-236, 1992.

[18] M. J. F. Gales and S. J. Young, "A Fast and Flexible Implementation of Parallel Model Combination," Proc. ICASSP95, pp. 133-136, 1995.

[19] M. J. F. Gales, "Model-Based Techniques for Noise Robust Speech Recognition," Dissertation submitted to the University of Cambridge for the degree of Doctor of Philosophy, Sep. 1995.

[20] F. Martin, K. Shikano, Y. Minami, Y. Okabe, "Recognition of Noisy Speech by Using the Composition of Hidden Markov Models," 日本音響学会講演論文, 1-7-10, pp. 65-66, Oct. 1992.

研究業績

[1] 坂井 伸圭, 中井 満, 下平 博, 嵯峨山 茂樹, “ヤコビ適応法を用いた雑音環境・伝達特性・話者への同時適応,” 日本音響学会 2001 年秋季研究発表会講演論文集, Oct. 2001.

[2] 坂井 伸圭, 中井 満, 下平 博, 嵯峨山 茂樹, “発声変形に対するヤコビ適応法,” 日本音響学会 2002 年春季研究発表会講演論文集, Mar. 2002.