音声言語情報処理 20−15
(1998. 2. 6)

雑音・残響環境下でのHMM分解・合成法によるハンズフリー音声認識

滝口 哲也　　中村 哲　　鹿野 清宏

奈良先端科学技術大学院大学 情報科学研究科
〒630-0101 奈良県生駒市高山町 8916-5
E-mail: tetuy-t@is.aist-nara.ac.jp

あらまし ユーザがマイクロホンから離れて発話した場合のハンズフリー音声認識では，周囲の雑音及び残響の影響を受けてしまい，認識精度が劣化してしまう．それらの影響に対処するために，本稿では，観測信号に対する尤度最大化にもとづいた，音響伝達特性HMMの推定方法を提案する．提案するHMM分解法は，雑音・残響環境下では2回適用される．まず，周波数領域において，雑音HMMからの分解をML推定にもとづいて行なう．更に，領域変換を行ない，ケプストラム領域において，音響伝達特性HMMをML推定にもとづいて分解する．また，この領域変換の際には，特徴パラメータを直接取り扱うのではなく，モデルパラメータを用いる．音素を単位にした500単語認識実験の結果，特定話者認識率が77.2%から91.2%に，不特定話者認識率は54.4%から66.2%に改善され，提案方法の有効性が示された．

キーワード ハンズフリー音声認識、残響、雑音、モデル適応、HMM分解

Hands-free Speech Recognition by HMM De-Composition in Noisy Reverberant Environments

Tetsuya Takiguchi, Satoshi Nakamura and Kiyohiro Shikano

Graduate School of Information Science, Nara Institute of Science and Technology, NAIST.
8916-5, Takayama-cho, Ikoma-shi, Nara, 630-0101, JAPAN
E-mail: tetuy-t@is.aist-nara.ac.jp

Abstract This paper proposes a new method to estimate HMM parameters of an acoustical transfer function based on HMM decomposition for hands-free speech recognition. This method is able to estimate the model parameters by maximizing a likelihood(ML) of noisy reverberant speech data in the model domain. The proposed HMM decomposition method is applied twice to noisy reverberant speech. Firstly, the HMM decomposition method is applied in the liner spectral domain to estimate the distorted speech HMMs by ML estimation. The obtained distorted speech HMMs are converted to the cepstral domain. Then the HMM decomposition method is applied again in the cepstral domain to estimate the acoustical transfer function HMM by ML estimation. The speaker dependent and independent recognition rates for distant-talking 500 words are improved from 77.2% to 91.2% and from 54.4% to 66.2%, respectively.

key words hands-free speech recognition, reverberation, noise, model adaptation, HMM decomposition

1 まえがき

現在の音声認識システムでは，ユーザは，マイクロホンの位置を意識しなくてはならない．なぜなら，ユーザがマイクロホンから離れて発話すると，マイクロホンへの入力音声は，周囲の雑音及び，残響の影響を受けてしまい，学習データと観測データとの間のミスマッチにより，認識率の劣化が生じてしまうためである．

これまでに，加法性雑音や電話回線などの乗法性歪みの要因に対処するための研究については，多く行なわれてきている．それらの研究は，音声強調による方法とモデル適応化による影響補償の方法に大別できる．音声強調としては，加法性雑音に対して Spectral Subtraction[3]，電話回線などの歪みに対して Cepstral Mean Subtraction [4] が提案されている．モデル適応化としては，加法性雑音に対して HMM 合成法 [5]，PMC[6]，電話回線などの歪みに対して Stochastic Matching[8] があげられる．これらは，加法性雑音または，電話回線等の歪みのどちらかに対処する方法であるが，両方の要因を取り扱う研究についても行なわれてきている [7][9]．文献 [9] では，対象とする環境モデルの定式化及びその推定アルゴリズムが，本手法と異なっているが．加法性雑音と電話回線やマイクロホンによる歪みの影響を受けた音声に対して，従来の HMM 合成法 [5][6] による合成 HMM の尤度を最大化することにより，電話回線歪みの推定を行なう方法を提案している．

著者らはこれまでに，HMM 合成を加法性雑音及び残響による影響を受けた音声の認識へ適用してきた [1][2]．合成 HMM は，もし信号源が互いに独立であるならば，作成することができる．音声と加法性雑音は，周波数領域では独立であり加法性が仮定されている．一方，音声と音響伝達特性は，周波数領域では積によって関係づけられているので，ケプストラム領域では独立であり加法性が仮定されている．ゆえに，雑音及び残響環境下において，HMM 合成法の適用が可能となる．著者らは，文献 [2] において，実際の環境における，HMM 合成法の有効性を認識実験によって示しているが，音響伝達特性 HMM のパラメータの推定方法についての問題が残されている．文献 [2] では，実際の部屋でインパルス応答を測定して，その値を利用して音響伝達特性 HMM のパラメータを求めている．つまり，認識を行なう前に，あらかじめ音響伝達特性を測定しておく必要がある．

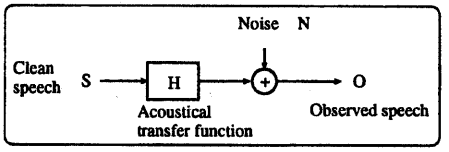

図 1: 対象とする環境のモデル

本稿では，HMM 分解に基づく音響伝達特性 HMM の推定方法を提案する．雑音・残響環境下では，この HMM 分解は 2 回（周波数領域とケプストラム領域）適用され，また領域変換の際には，特徴パラメータを直接取り扱うのではなく，その統計量が用いられる．この方法では，あらかじめインパルス応答を測定しておく必要はなく，更に，発話者の位置が既知である必要もなく，任意の場所から発話された音声を用いて，その場所からマイクロホンへの音響伝達特性を推定する．また，そのように推定された音響伝達特性を基に，いくつかの代表的な音響伝達特性 HMM を用意することで，そのエルゴディック HMM により移動音源に対する認識 [1][2] が可能となる．

2 雑音・残響環境下での音声認識法

まず，HMM 合成法による，雑音・残響環境下での音声認識方法について説明する [1][2]．

雑音・残響環境下でのモデルは，図 1 のように表される．この時，観測信号は

$$O(\omega;m) = S(\omega;m) \cdot H(\omega;m) + N(\omega;m) \quad (1)$$

と表される．ここで $O(\omega;m)$,$S(\omega;m)$,$H(\omega;m)$, $N(\omega;m)$ は，各々，フレーム番号 m，周波数 ω の観測信号，クリーン音声，音響伝達特性，雑音のパワースペクトルを表している．HMM 合成法は，加算条件の成立する領域において適用されるので，式 (1) を，次のように書きかえる．

$$O(\omega;m) = \exp\{\mathcal{F}(S_{cep}(t;m) + H_{cep}(t;m))\} + N(\omega;m) \quad (2)$$

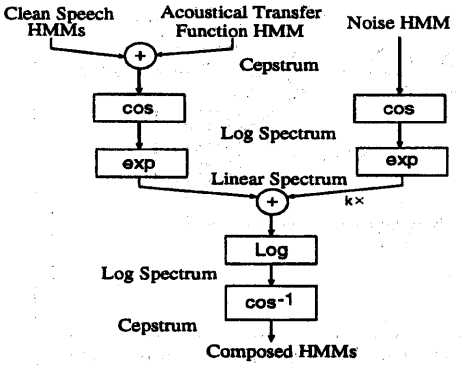

図 2: 出力確率の合成アルゴリズム

ここで，$S_{cep}(t;m)$，$H_{cep}(t;m)$ は，各々，クリーン音声，音響伝達特性のフレーム番号 m，ケフレンシー t におけるケプストラムを表している．exp，$\mathcal{F}$ は，各々，指数変換，コサイン変換を表す．従って，合成 HMM の出力確率分布は，式 (2) をモデル領域において適用することにより，求めることができる．図 2 にそのアルゴリズムを示す．また，合成 HMM の状態数及び遷移確率などは，各々の積により求めることができる．

ここで，HMM 合成法を適用するためには，まず，各々のモデル（ここでは，クリーン音声，雑音と音響伝達特性）を作成しておく必要がある．雑音 HMM は，観測信号の無音区間より推定を行なう．また，音響伝達特性の推定方法については，以下で説明する．

3 ML 推定に基づく HMM 分解によるパラメータ推定

観測データに対する合成モデルの尤度が最大になるようにして，音響伝達特性 HMM を求める．

$$\hat{\lambda}_H = \underset{\lambda_H}{\operatorname{argmax}} P(O|\lambda_H, \lambda_N, \lambda_S)$$

ここで，λ はモデルパラメータの集合を表す．観測データより音響伝達特性 HMM を推定する方法には，モデルの合成の逆のプロセスであるモデルの分解を用いる．この時，一つのモデルを既知として，もう一つのモデルとの分解を行なう．電話回線歪みの推定を EM アルゴリズムにより行なう方法については，文献 [8] で入力パラメータを直接用いる方法が提案されている．しかしながら，加法性雑音と電話回線歪みなどの乗法性歪みが同時に存在する場合，特徴パラメータに対して線形・非線形の領域変換を行なう必要があり，入力データが長くなると計算量の増加が問題となる．一方，HMM 分解法では，特徴パラメータを直接取り扱うのではなく，モデルパラメータであるその統計量を用いるので，少量のモデルパラメータの領域変換を行なうだけでよい．

式 (2) より，音響伝達特性 HMM は，次のように表される．

$$\lambda_{H_{cep}} = \mathcal{F}^{-1}\{\log(\lambda_{O_{lin}} \ominus \lambda_{N_{lin}})\} \ominus \lambda_{S_{cep}}$$

ここで添字の cep と lin は各々ケプストラム領域とパワースペクトラム領域を表している．また，$\ominus$ は HMM の分解を表す．このように，雑音・残響環境下では，HMM 分解は 2 回適用される．まず，パワースペクトラム領域において，雑音 HMM を固定して，Distorted Speech HMMs を最尤推定にもとづいて求める（3.1節）．更に，Distorted Speech HMMs をケプストラム領域へ変換し，Clean Speech HMMs を固定して，音響伝達特性 HMM を最尤推定にもとづいて求める（3.2節）．以下に，このモデル適応化アルゴリズムを示す．

1. 雑音及び残響環境下で観測された適応データを用いて ML (Maximum Likelihood) 推定により $\lambda_{O_{cep}}$ のパラメータ推定を行なう．また，雑音 HMM $\lambda_{N_{cep}}$ を無音区間より推定し，各々をパワースペクトラム領域に変換する（$\lambda_{O_{lin}}, \lambda_{N_{lin}}$）．それから，$\lambda_{O_{lin}}$ から $\lambda_{SH_{lin}}$ を分解する．

$$\begin{aligned}\hat{\lambda}_{SH_{lin}} &= \underset{\lambda_{SH_{lin}}}{\operatorname{argmax}} P(SH+N|\lambda_{SH_{lin}}, \lambda_{N_{lin}}) \\ &\triangleq \lambda_{O_{lin}} \ominus \lambda_{N_{lin}}\end{aligned}$$

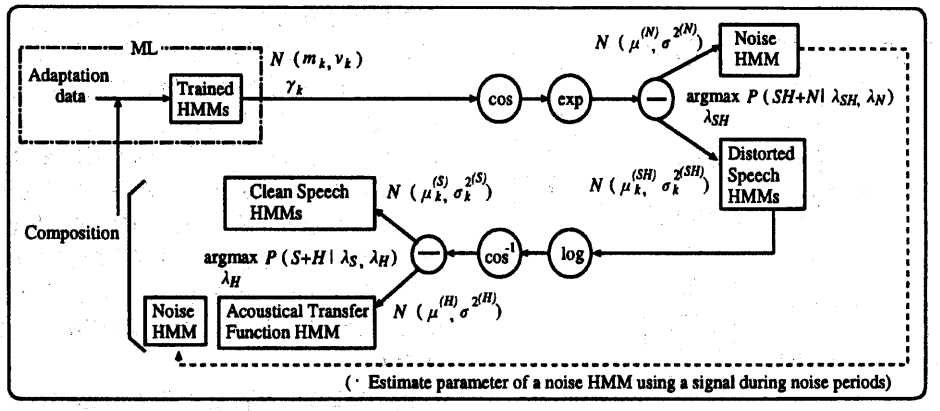

図 3: HMM 分解によるパラメータ推定法

ここで，$SH+N$ は，式 (1) で表されるものとする．

2. $\hat{\lambda}_{SH_{lin}}$ をケプストラム領域へ変換し，$\lambda_{(S+H)_{cep}}$ から $\lambda_{H_{cep}}$ を分解する．

$$\hat{\lambda}_{H_{cep}} = \underset{\lambda_{H_{cep}}}{\operatorname{argmax}} P(S+H|\lambda_{H_{cep}}, \lambda_{S_{cep}})$$
$$\triangleq \lambda_{(S+H)_{cep}} \ominus \lambda_{S_{cep}}$$

ここで，$S+H$ は，式 (2) で示されているように，音声と音響伝達特性のケプストラム上での関係を表す．

以上の手続きを図 3 に示す．出力確率分布は，混合数 K の Tied Mixture 型分布とし，ML の一回の推定により得られた適応データの平均値と分散を，μ_k と v_k とする．この分布をパワースペクトラム領域へ変換し，雑音との分解を行なう．次に，ケプストラム領域へ変換し，クリーン音声 HMM との分解を行ない，音響伝達特性 HMM を推定する．更に，クリーン音声 HMM と雑音，伝達特性の合成 HMM を作成する．この手続きを，尤度が収束するまで繰り返し行なう．

3.1 雑音 HMM からの分解

モデルパラメータ $\lambda_{O_{cep}}$ をパワースペクトラム領域に変換し，Distorted Speech HMMs ($\lambda_{SH_{lin}}$) の分解を行なう．

$$\hat{\lambda}_{SH_{lin}} = \underset{\lambda_{SH_{lin}}}{\operatorname{argmax}} P(SH+N|\lambda_{SH_{lin}}, \lambda_{N_{lin}})$$

これを EM アルゴリズムにより解く．ここで，$SH+N$ は，式 (1) で表されるものとする．まず，Q 関数を以下のように定義する．(以下では，添字の lin を省略する)

$$Q(\hat{\lambda}_{SH}|\lambda_{SH}) = \sum_{p=1}^{P}\sum_{n=1}^{W_p}\sum_{s^{(p,n)}}\sum_{k^{(p,n)}} \frac{f(o^{(p,n)}, s^{(p,n)}, k^{(p,n)}|\lambda_{SH}, \lambda_N)}{f(o^{(p,n)}|\lambda_{SH}, \lambda_N)} \times \log f(o^{(p,n)}, s^{(p,n)}, k^{(p,n)}|\hat{\lambda}_{SH}, \lambda_N)$$

$$f(o, s, k|\lambda_{SH}, \lambda_N) = \prod_{t=1}^{T}\{a_{s_{t-1}s_t}\omega_{s_t,k_t} N(o_t; \mu_{k_t}^{(SH)} + \mu^{(N)}, \sigma_{k_t}^{2(SH)} + \sigma^{2(N)})$$

ここで，P は音韻数で，それぞれの音韻は，W_p 個の適応データを持つとする．また，$o^{(p,n)}$ は，音韻 p に関連する n 番目の観測系列で，長さ $T^{(p,n)}$ とし，$s^{(p,n)}$, $k^{(p,n)}$ は，各々，$o^{(p,n)}$ に対する状態系列，混合要素の系列とする．また，λ_{SH} の出力確率分布を混合数 K，次元数 D の平均 $\mu_k^{(SH)}$，分散 $\sigma_k^{2(SH)}$ の Tied Mixture 型分

布とし，合成HMMの出力確率分布に対する重みを $\omega_{s,k}$ とする．また，λ_N の出力確率分布を平均 $\mu^{(N)}$，分散 $\sigma^{2(N)}$ の単一ガウス分布とする．いま，Q 関数の出力確率分布に関する項 $Q_{\hat{\theta}_k}$ ($\hat{\theta}_k = \{\hat{\mu}_k^{(SH)}, \hat{\sigma}_k^{2(SH)}\}$) に注目すると，

$$
\begin{aligned}
&Q_{\hat{\theta}_k}(\hat{\lambda}_{SH}|\lambda_{SH}) \\
&= \sum_{p=1}^{P}\sum_{n=1}^{W_p}\sum_{t=1}^{T^{(p,n)}} \gamma_{t,k}^{(p,n)} \\
&\quad \times \log N(o_t^{(p,n)}; \hat{\mu}_k^{(SH)} + \mu^{(N)},\ \hat{\sigma}_k^{2(SH)} + \sigma^{2(N)}) \\
&= -\sum_{p=1}^{P}\sum_{n=1}^{W_p}\sum_{t=1}^{T^{(p,n)}} \gamma_{t,k}^{(p,n)} \\
&\qquad \times \left\{ \frac{1}{2}\log(2\pi)^D(\hat{\sigma}_k^{2(SH)} + \sigma^{2(N)}) + \right. \\
&\left. \frac{(o_t^{(p,n)} - \hat{\mu}_k^{(SH)} - \mu^{(N)})'(o_t^{(p,n)} - \hat{\mu}_k^{(SH)} - \mu^{(N)})}{2(\hat{\sigma}_k^{2(SH)} + \sigma^{2(N)})} \right\}
\end{aligned}
$$

この $Q_{\hat{\theta}_k}$ 関数を最大にする $\hat{\lambda}_{SH}$ は，$\hat{\lambda}_{SH}$ に関する偏微分により求まる．ここで，$'$ は転置を表す．

$$
\hat{\mu}_k^{(SH)} = m_k - \mu^{(N)}, \qquad \hat{\sigma}_k^{2(SH)} = v_k - \sigma^{2(N)}
$$

ここで，

$$
m_k = \sum_p\sum_n\sum_t \gamma_{t,k}^{(p,n)} o_t^{(p,n)} / \gamma_k
$$

$$
v_k = \sum_p\sum_n\sum_t \gamma_{t,k}^{(p,n)} (o_t^{(p,n)} - m_k)^2 / \gamma_k
$$

である．

3.2 音響伝達特性HMMの分解

モデルパラメータ $\lambda_{SH_{lin}}$ をケプストラム領域に変換し，音響伝達特性HMMの分解を行なう．

$$
\hat{\lambda}_{H_{cep}} = \underset{\lambda_{H_{cep}}}{\operatorname{argmax}} P(S+H|\lambda_{H_{cep}}, \lambda_{S_{cep}})
$$

これをEMアルゴリズムにより解く．ここで，$S+H$ は，式(2)で示されているように，音声と音響伝達特性のケプストラム上での関係を表す．3.1節と同様にして $Q_{\hat{\theta}}$ 関数 ($\hat{\theta} = \{\hat{\mu}^{(H)}, \hat{\sigma}^{2(H)}\}$) を以下のように定義する．ここで，$\lambda_S$ の出力確率分布を混合数 K，平均 $\mu_k^{(S)}$，分散 $\sigma_k^{2(S)}$ のTied Mixture型分布とし，λ_H の出力確率分布を平均 $\mu^{(H)}$，分散 $\sigma^{2(H)}$ の単一ガウス分布とする．

$$
\begin{aligned}
&Q_{\hat{\theta}}(\hat{\lambda}_H|\lambda_H) \\
&= \sum_{p=1}^{P}\sum_{n=1}^{W_p}\sum_{t=1}^{T^{(p,n)}}\sum_{k}^{K} \gamma_{t,k}^{(p,n)} \\
&\quad \times \log N(o_t^{(p,n)}; \mu_k^{(S)} + \hat{\mu}^{(H)},\ \sigma_k^{2(S)} + \hat{\sigma}^{2(H)}) \\
&= -\sum_{p=1}^{P}\sum_{n=1}^{W_p}\sum_{t=1}^{T^{(p,n)}}\sum_{k}^{K} \gamma_{t,k}^{(p,n)} \\
&\qquad \times \left\{ \frac{1}{2}\log(2\pi)^D(\sigma_k^{2(S)} + \hat{\sigma}^{2(H)}) + \right. \\
&\left. \frac{(o_t^{(p,n)} - \mu_k^{(S)} - \hat{\mu}^{(H)})'(o_t^{(p,n)} - \mu_k^{(S)} - \hat{\mu}^{(H)})}{2(\sigma_k^{2(S)} + \hat{\sigma}^{2(H)})} \right\}
\end{aligned}
$$

ここでは，3.1節とは異なり（更に混合数 k に関する和をとるため），$\mu^{(H)}$ と $\sigma^{2(H)}$ に関する偏微分を求めるのは，困難である．そこで，以下のように，音響伝達特性の変化量を Δ で表し，$\Delta\hat{\mu}^{(H)}$ と $\Delta\hat{\sigma}^{2(H)}$ に関する偏微分により，推定式を求める．

$$
\hat{\mu}^{(H)} = \mu^{(H)} + \Delta\hat{\mu}^{(H)}, \quad \hat{\sigma}^{2(H)} = \sigma^{2(H)} + \Delta\hat{\sigma}^{2(H)}
$$

まず，音響伝達特性 H ($\Delta\hat{\mu}^{(H)}$) の再推定式に関しては，$\partial Q_{\hat{\theta}}(\hat{\lambda}_H|\lambda_H)/\partial\Delta\hat{\mu}^{(H)} = 0$ より，

$$
\Delta\hat{\mu}^{(H)} = \frac{\displaystyle\sum_{k=1}^{K} \gamma_k \frac{\hat{\mu}_k^{(SH)} - \mu_k^{(S)} - \mu^{(H)}}{\sigma_k^{2(S)} + \sigma^{2(H)}}}{\displaystyle\sum_{k=1}^{K} \frac{\gamma_k}{\sigma_k^{2(S)} + \sigma^{2(H)}}}
$$

となる．また，分散に関しては，$\partial Q_{\hat{\theta}}(\hat{\lambda}_H|\lambda_H)/\partial\Delta\hat{\sigma}^{2(H)} = 0$ より，

$$
\sum_{k}^{K} \gamma_k \left\{ \frac{\sigma_k^{2(S)} + \sigma^{2(H)} + \Delta\sigma^{2(H)} - \phi_k}{(\sigma_k^{2(S)} + \sigma^{2(H)} + \Delta\sigma^{2(H)})^2} \right\} = 0 \quad (3)
$$

ここで，

$$
\begin{aligned}
\phi_k &= \sigma_k^{2(SH)} + \mu_k^{2(SH)} \\
&\quad + (\mu_k^{(S)} + \hat{\mu}^{(H)})(\mu_k^{(S)} + \hat{\mu}^{(H)} - 2\mu_k^{(SH)})
\end{aligned}
$$

とする．ところで，式(3)の $\{\}$ の箇所に注目し，以下のような関数 F を定義すると，

$$
F(\Delta\sigma^{2(H)}) = \frac{\sigma_k^{2(S)} + \sigma^{2(H)} + \Delta\sigma^{2(H)} - \phi_k}{(\sigma_k^{2(S)} + \sigma^{2(H)} + \Delta\sigma^{2(H)})^2}
$$

$\Delta\sigma^{2(H)}$ は EM アルゴリズムにより，十分小さく 0 に収束する値なので，この式の原点におけるテイラー展開を行ない，1 次の項までを用い，次式を得る．

$$\Delta\hat{\sigma}^{2(H)} \simeq \frac{\sum_k^K \gamma_k \left\{ \frac{1}{\sigma_k^{2(S)} + \sigma^{2(H)}} - \frac{\phi_k}{(\sigma_k^{2(S)} + \sigma^{2(H)})^2} \right\}}{\sum_k^K \gamma_k \left\{ \frac{1}{(\sigma_k^{2(S)} + \sigma^{2(H)})^2} - \frac{2\phi_k}{(\sigma_k^{2(S)} + \sigma^{2(H)})^3} \right\}}$$

4 認識実験

ここでは，以下の実験を行なう．

- シミュレーション評価：
  残響のみの場合（Noise-free）
- 実環境下での評価

4.1 実験条件

実験で使用した特定話者モデル（54 音韻）は，2620 単語より学習されている．不特定話者モデルは，ASJ の連続音声データベースの 64 人分約 9600 文章より学習されている．ここで学習データにおいて，スピーカ特性を補正するために，無響室で測定されたスピーカ特性を各々のデータベースのクリーン音声に波形上で畳み込んでおく．クリーン音声 HMM は，3 状態 3 ループの 256 混合 Tied Mixture 型対角共分散 HMM である．雑音 HMM と音響伝達特性 HMM は，各々 1 状態，単一ガウス分布とする．評価用データとして学習で使用していない 500 単語を使用し，適応データとして音素バランス単語を 3 種類の集合に分けて使用する．特定話者認識実験では，ATR 音声データベースより男性 1 名を，不特定話者認識実験では，男性 2 名，女性 1 名を評価データとして使用する．

- シミュレーションデータ
  残響だけの環境下 (Noise-free) における，HMM 分解法の評価を行なうために，図 4 に示される簡易音響実験室（残響時間 約 180msec）において，各音源位置からマイクロホンへの音響伝達特性の測定を行なう．それらの測定された音響伝達特性を ATR 音声データベースのクリーン音声と波形上で畳み込んでテストデータと適応データを作成する．

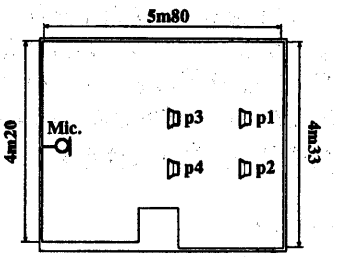

図 4: 簡易音響実験室

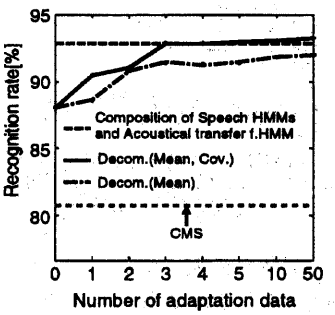

図 5: 残響環境下（Noise-free）での認識結果

- 実環境データ
  図 4の各音源位置から ATR 音声データベースのクリーン音声を収録する．背景雑音は，換気扇，計算機雑音等で，平均 SNR は，16.7dB である．

4.2 実験結果

4.2.1 シミュレーション評価結果

まず，シミュレーション実験で，残響環境下の認識を行なう．図 5 に特定話者に対する場所 4 箇所の平均認識率を示す．HMM 分解法により，音響伝達特性を推定し，HMM 合成法によりクリーン音声 HMM と合成し，認識した結果を “Decom.(Mean)” と”Decom.(Mean,Cov.)

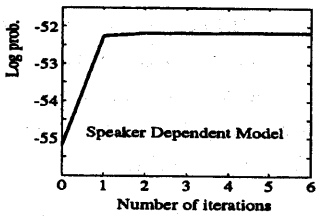

図 6: HMM 分解法の収束性

" に示す．ここで,"Decom. (Mean)" と"Decom. (Mean,Cov.) " は，各々，平均値のみ，平均値と分散の適応を表す．クリーン音声 HMM での平均認識率は，88.1%で，提案手法を用いることにより，平均値のみの適応で，91.8%（適応データ数：10 単語）まで認識率が改善されている．また，分散も適応することにより，約 1%程度の認識率の改善が得られている．次に，文献 [1][2] で提案した方法で，音響伝達特性が既知の場合の合成 HMM による認識結果について示す．この時，音響伝達特性 HMM の平均値 $\mu^{(H)}$ は，以下のようにして求める．

$$\mu^{(H)} = \frac{1}{g}\sum_{j=1}^{g}(c'(j) - c(j)) = \frac{1}{g}\sum_{j=1}^{g}h(j)$$

ここで，g は音響伝達特性を求めるために使用する学習データの全フレーム数である（ここでは，HMM 学習の際に使用した 500 単語を用いた）．$c'(j)$ と $c(j)$ は，各々，音響伝達特性が畳み込まれた音声と，畳み込まれる前のクリーン音声の j 番目のフレームのケプストラムである．また，分散 $\sigma^{2(H)}$ は，

$$\sigma^{2(H)} = \frac{1}{g}\sum_{j=1}^{g}(h(j) - \mu^{(H)})(h(j) - \mu^{(H)})$$

とする．図5 より，HMM 分解法により音響伝達特性を推定した場合の認識率が，音響伝達特性が既知の場合の認識率に近付いているのが分かる．また,"CMS"（Cepstral Mean Subtraction）との比較実験では，残響時間が 180msec の場合，あまり効果がないのがわかる．ここでの"CMS" は，一単語毎にケプストラム平均値を計算している．

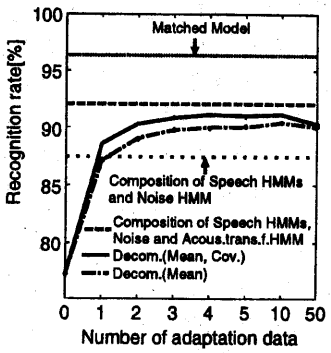

図 7: 実環境下での認識結果（特定話者認識）

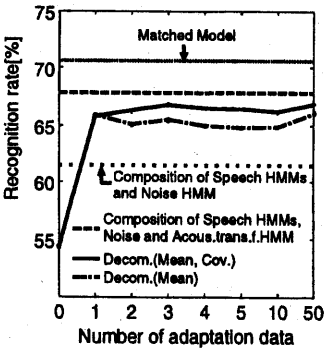

図 8: 実環境下での認識結果（不特定話者認識）

図 6 に平均対数尤度とアルゴリズムの反復回数を示す．HMM 分解法では，3 回くらいの繰り返しで尤度が収束しているのが分かる．

4.2.2 実環境下での評価結果

次に，実環境下での認識を行なう．図 7，図 8 に特定話者と不特定話者に対する場所 4 箇所の平均認識率を示す．クリーン音声 HMM での平均認識率は，特定，不特定に対して，各々，77.2%，54.4%であり，また，雑音 HMM とクリーン音声 HMM の合成 HMM で認識した場合の平均認識率は，各々，87.5%，61.5% である．HMM 分解法により，音響伝達特性を推定することにより，認識率は，平均値のみの適応で，各々，90.5%，64.9%（適応データ数：10 単語）

まで改善されている．また，分散も適応することにより，各々，91.2%，66.2%まで改善されている．音響伝達特性を既知とした場合 [1][2] の結果と比べると，提案手法は，その認識率に近付いているのがわかる．“Matched Model” は，測定した音響伝達特性をクリーン音声と波形上で畳み込み，更に，背景雑音を計算機上で付加したデータ（特定 2620 単語，不特定約 9600 文章）で学習したモデルで認識した場合の結果である．この結果（不特定に対しては，場所 p1 のみ）と比べると HMM 分解による推定精度は，特定，不特定において，各々 5.2%，4.5%の改善の余地が残されている．

5 むすび

本稿では，HMM 分解・合成法による雑音及び残響環境下でのモデル適応化法を提案した．HMM 分解法により，あらかじめ音響伝達特性を測定しておく必要はなくなり，また，発話者の位置が既知である必要もなく，任意の場所から発話された音声を用いて，音響伝達特性の推定が可能となる．

実環境下での評価実験では，音響伝達特性が既知として HMM 合成法により認識した場合の結果に [1][2]，提案手法により音響伝達特性を推定した場合の結果が近付いていることがわかり，提案手法の有効性が示せた．しかし，音響伝達特性の推定精度は，まだ不十分であり，今後は，残響の影響がフレーム間に渡ってしまうことに対処する方法について検討していく．

参考文献

[1] 滝口 哲也, 中村 哲, 鹿野 清宏, “加法性雑音，伝達特性による歪みを受けた音声の HMM 合成による認識”, 音響講論集, 1-2-2, pp.3-4, 1995.

[2] 滝口 哲也, 中村 哲, 鹿野 清宏, “雑音と残響のある環境下での HMM 合成によるハンズフリー音声認識法”, 信学論（D-II）, vol.J79-D-II, No.12, pp.2047-2053, Dec.1996.

[3] S.F.Boll. “Suppression of Acoustic Noise in Speech Using Spectral Subtraction”, IEEE Trans,ASSP-27,pp.113-120,1979.

[4] Atal.B, “Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification”, Proc. J.Acoust.Soc.Amer., Vol.55, pp.1304-1312,1974.

[5] F.Martin, K.Shikano and Y.Okabe, “Recognition of Noisy Speech by Composition of Hidden Markov Models”, 信学技報, SP92-96, 1992.

[6] M.J.F.Gales and F.J.Young, “An improved Approach to the Hidden Markov Model Decomposition of Speech and Noise”, Proc.ICASSP, pp.233-236,1992.

[7] M.J.F.Gales and S.J.Young, “PMC for Speech Recognition in Additive and Convolutional Noise”, CUED-F-INFENG-TR154, 12, 1993.

[8] A.Sankar and C-H.Lee, “Robust Speech Recognition Based on Stochastic Matching”, Proc. ICASSP, pp.121-124,1995.

[9] 南 泰浩, 古井 貞熙, “HMM 合成に基づく尤度最大化適応法”, 信学技報, SP95-24, pp.45-50, 6, 1995.