論　文

# 雑音・残響環境下でのHMM分解・合成法によるモデル適応化

滝口　哲也[†]　　中村　哲[†]　　鹿野　清宏[†]

# Model Adaptation by HMM Decomposition and Composition in Noisy Reverberant Environments

Tetsuya TAKIGUCHI[†], Satoshi NAKAMURA[†], and Kiyohiro SHIKANO[†]

**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に対しては，残響環境下において認識精度が劣化してしまう．なぜなら，その音声は，周囲の雑音およ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にミスマッチが生じてしまうためである．それらの影響に対処するために，筆者らは，これまでに音響伝達特性HMMを作成し，HMM合成法による音声認識法を提案した[1],[2]．しかし，その方法では認識を行う前に，あらかじめ各場所からの音響伝達特性を測定する必要があった．本論文では，音響伝達特性HMMの推定を，観測信号より行う方法を提案する．この方法では，話者の場所が既知である必要はなく，任意の場所から発話された適応データを用いて，最ゆう推定に基づき，あるHMMを一つの既知HMMともう一つのHMMに分解し，モデルパラメータの推定を行う．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**　雑音，残響，モデル適応，HMM分解・合成，ハンズフリー音声認識

## 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およ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ひずみの要因に対処するための研究については，多く行われてきている．それらの研究は，音声強調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]，電話回線などのひずみに対して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]，電話回線などのひずみに対して，時系列パラメータを直接用いてひずみの推定を行うStochastic Matching [8],[9]が挙げられる．これらは，加法性雑音または，電話回線等のひず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も行われてきている[7],[10]〜[13]．文献[10]では，加法性雑音と電話回線やマイクロホンによるひず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のゆう度を最大化することにより，電話回線ひずみの推定を行う方法を提案しているが，対象とする環境モデルが残響環境下とは異なる．また文献[13]でも，文献[10]と同様の環境モデルを用いて乗法性ひずみと加法性雑音への適応を試みているが，文献[8],[9]の方法に基づいているため，適応には観測値の時系列パラメータを直接用いる必要がある．一方，HMM分解法[11],[12]では，加法性雑音と乗法性ひずみが存在する環境下で，2段階（パワースペクトラム領域とケプストラム領域）の最ゆう推定を用いて乗法性ひずみの推定を行う．更に，このHMM分解法の特徴は，観測値の時系列パラメータを用いるのではなく，観測値の統計量を用いてモデルパラメータの推定を行う点にある．

筆者らはこれまでに，HMM合成を加法性雑音およ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

[†]奈良先端科学技術大学院大学情報科学研究科，生駒市
[†]Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma-shi, 630-0101 Japan

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．故に，雑音および残響環境下において，HMM 合成法の適用が可能となる．筆者らは文献[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た．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った．

本論文では，HMM 分解に基づく音響伝達特性 HMM の推定方法を提案する．提案方法は，HMM 合成の逆の過程により導かれる．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もとに，いくつかの代表的な音響伝達特性 HMM を用意し，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．このとき，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．$O(\omega;m)$，$S(\omega;m)$，$H(\omega;m)$，$N(\omega;m)$ は，各々，フレーム番号 $m$，周波数 $\omega$ 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) を次のように書き換える．

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

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，cos 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 の状態数および遷移確率などは，各々の積により求めることができる．

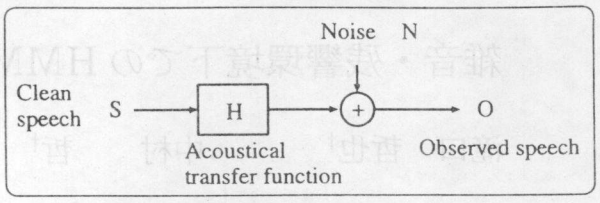


図 1 対象とする環境のモデル
Fig. 1 Environment model.

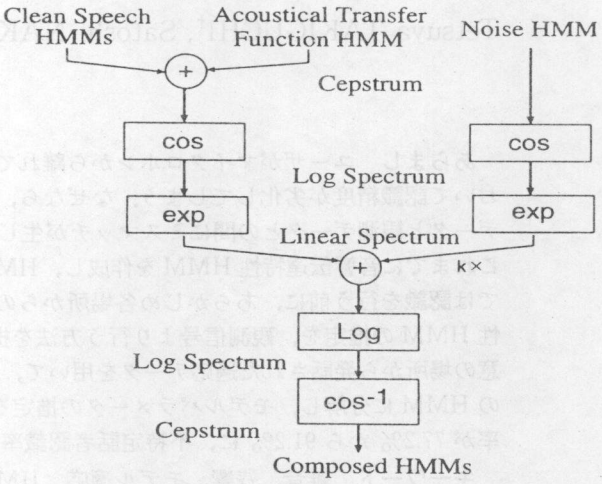


図 2 出力確率の合成アルゴリズム
Fig. 2 Block diagram of HMM composition.

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

### 3.1 原　　理

観測データに対する合成モデルのゆう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，$\lambda$ 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，

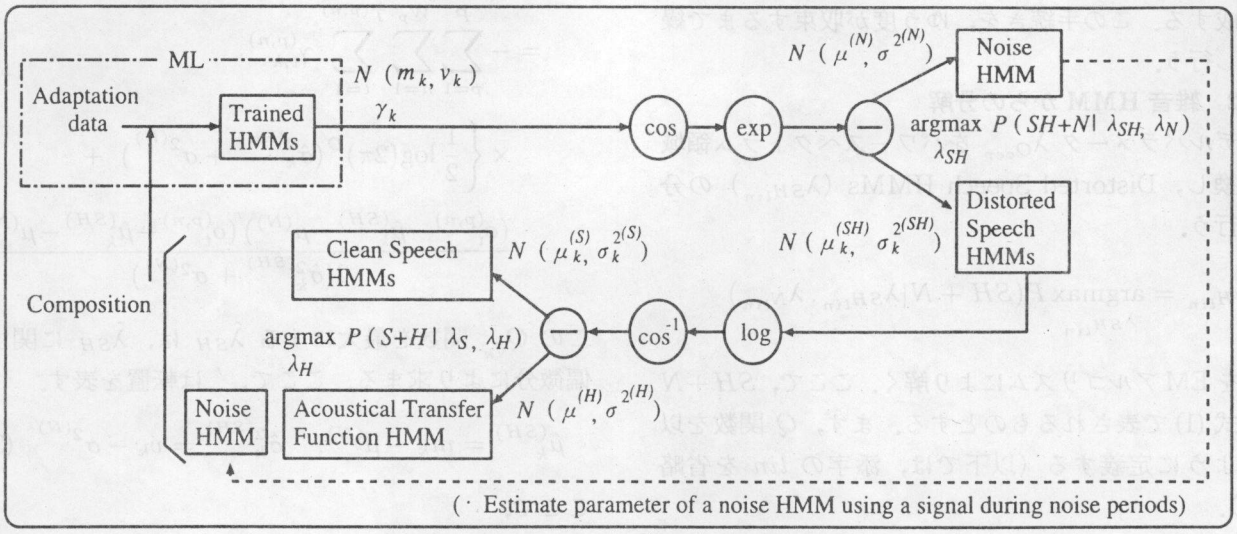


図 3　HMM 分解によるパラメータ推定法
Fig. 3　Parameter estimation by HMM decomposition.

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．このとき，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う．電話回線ひずみの推定を EM アルゴリズムにより行う方法については，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ひずみなどの乗法性ひず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．

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

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 を最ゆう推定に基づいて求める (3.2)．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 を最ゆう推定に基づいて求める (3.3)．以下に，このモデル適応化アルゴリズムを示す．

(1)　雑音および残響環境下で観測された適応データ（発話内容は既知）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．

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

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．初期 HMM は，クリーン音声 HMM と雑音 HMM の合成 HMM とする．また，出力確率分布は，混合数 $K$ の Tied Mixture とし，ML の 1 回の推定により得られた適応データの平均値と分散を，$m_k$ 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，ゆう度が収束するまで繰り返し行う．

### 3.2 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）．

$$\begin{aligned}
&Q(\hat{\lambda}_{SH}|\lambda_{SH}) \\
&= \sum_{p=1}^{P}\sum_{n=1}^{W_p}\sum_{s^{(p,n)}}\sum_{k^{(p,n)}} \\
&\quad \frac{f(o^{(p,n)}, s^{(p,n)}, k^{(p,n)}|\lambda_{SH}, \lambda_N)}{f(o^{(p,n)}|\lambda_{SH}, \lambda_N)} \\
&\quad \times \log f(o^{(p,n)}, s^{(p,n)}, k^{(p,n)}|\hat{\lambda}_{SH}, \lambda_N) \\
&f(o, s, k|\lambda_{SH}, \lambda_N) \\
&= \prod_{t=1}^{T}\{a_{s_{t-1}s_t}\omega_{s_t,k_t} \\
&\quad N(o_t; \mu_{k_t}^{(SH)} + \mu^{(N)},\ \sigma_{k_t}^{2(SH)} + \sigma^{2(N)})\}
\end{aligned}$$

ここで，$P$ は音韻数で，それぞれの音韻は，$W_p$ 個の適応データをもつとする．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，$\lambda_{SH}$ の出力確率分布を混合数 $K$，次元数 $D$ の平均 $\mu_k^{(SH)}$，分散 $\sigma_k^{2(SH)}$ （対角共分散行列）の Tied Mixture HMMs 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，$\lambda_N$ 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，

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

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．

$$\hat{\mu}_k^{(SH)} = m_k - \mu^{(N)}, \quad \hat{\sigma}_k^{2(SH)} = v_k - \sigma^{2(N)} \tag{3}$$

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．このように HMM 分解法では，雑音を除去したモデル $\lambda^{(SH)}$ を求める際に，観測値の統計量 $(m_k, v_k)$ を用いる．

### 3.3 クリーン音声 HMM からの分解

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，$O = S+H$ 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．しかし，このようなケプストラム時系列を実際に観測することはできない．そこで HMM 分解法では，時系列パラメータを直接用いるのではなく，3.2 で式 (3) より雑音 HMM からの分解によって求めた $S+H$ の統計量を用いて，更にクリーン音声 HMM からの分解を行う．式の導出における文献 [8], [9] との差は，$S+H$ の時系列パラメータを直接用いて乗法性ひずみの推定を行うのではなく，$S+H$ の統計量を用いて乗法性ひずみの推定を行う点にある．

3.2 と同様にして $Q_{\hat{\theta}}$ 関数 ($\hat{\theta} = \{\hat{\mu}^{(H)}, \hat{\sigma}^{2(H)}\}$) を式 (4) 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 HMMs とし，$\lambda_H$ 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．ここで，実際の観測信号は，

音声信号と音響伝達特性の波形上での畳込みによって表されるが，クリーン音声 HMM と音響伝達特性 HMM の合成 HMM では，フレーム単位（残響時間よりも短い）で処理を行うので，実際の信号の畳込みが正確には実現できない．従って，分散 $\sigma^{2(H)}$ を用いることによりこの差を補正する．また，雑音が理想的な加法性雑音である場合，雑音重畳音声モデルに対する雑音重畳音声データの状態アライメントと，雑音除去モデルに対する雑音除去音声データの状態アライメントはほぼ等しいと考えられる．そこでここでの分解の際にも，3.2 で用いた $\gamma$ と同じ値を用いるとする．

$$
\begin{aligned}
&Q_{\hat{\theta}}(\hat{\lambda}_H|\lambda_H) \\
&= \sum_{p=1}^{P}\sum_{n=1}^{W_p}\sum_{t=1}^{T^{(p,n)}}\sum_{k}^{K}\gamma_{t,k}^{(p,n)} \\
&\quad \times \log N(o_t^{(p,n)}; \mu_k^{(S)}+\hat{\mu}^{(H)},\ \sigma_k^{2(S)}+\hat{\sigma}^{2(H)}) \\
&= -\sum_{p=1}^{P}\sum_{n=1}^{W_p}\sum_{t=1}^{T^{(p,n)}}\sum_{k}^{K}\gamma_{t,k}^{(p,n)} \\
&\quad \times \left\{ \frac{1}{2}\log(2\pi)^D(\sigma_k^{2(S)}+\hat{\sigma}^{2(H)}) + \right. \\
&\quad \left. \frac{(o_t^{(p,n)}-\mu_k^{(S)}-\hat{\mu}^{(H)})'(o_t^{(p,n)}-\mu_k^{(S)}-\hat{\mu}^{(H)})}{2(\sigma_k^{2(S)}+\hat{\sigma}^{2(H)})} \right\}
\end{aligned}
\tag{4}
$$

ここでは，3.2 とは異なり（更に混合数 $k$ に関する和をとる），$\mu^{(H)}$ 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 $\Delta$ 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，

$$
\begin{aligned}
&\frac{\partial Q_{\hat{\theta}}(\hat{\lambda}_H|\lambda_H)}{\partial\Delta\hat{\mu}^{(H)}} \\
&= \sum_{p=1}^{P}\sum_{n=1}^{W_p}\sum_{t=1}^{T^{(p,n)}}\sum_{k}^{K} \\
&\quad \gamma_{t,k}^{(p,n)}\frac{o_t^{(p,n)}-\mu_k^{(S)}-\mu^{(H)}-\Delta\hat{\mu}^{(H)}}{\sigma_k^{2(S)}+\sigma^{2(H)}+\Delta\hat{\sigma}^{2(H)}} = 0
\end{aligned}
\tag{5}
$$

ここで，式 (3) により求めた雑音を分解したモデルの平均値 $\hat{\mu}_k^{(SH)}$ は，ケプストラム領域では以下のように表される．

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

$$
\hat{\sigma}_k^{2(SH)} = \sum_p\sum_n\sum_t \gamma_{t,k}^{(p,n)} (o_t^{(p,n)} - \hat{\mu}_k^{(SH)})^2 / \gamma_k
$$

ここで，この $o_t^{(p,n)}$ は雑音が除去された仮想的なケプストラム系列を表している．またこの出力確率分布 $N(\hat{\mu}_k^{(SH)}, \hat{\sigma}_k^{2(SH)})$ は，式 (3) より求めた雑音分解後の統計量 $N(m_k-\mu^{(N)}, v_k-\sigma^{2(N)})$ をケプストラム領域へ変換することにより既に求められている．

$$
\begin{aligned}
&\cos^{-1}[\ \log\{\ N(m_k-\mu^{(N)}, v_k-\sigma^{2(N)})\ \}\ ] \\
&\Rightarrow N(\hat{\mu}_k^{(SH)}, \hat{\sigma}_k^{2(SH)})
\end{aligned}
$$

従って，雑音が除去された仮想的なケプストラム時系列 $o_t^{(p,n)}$ の代わりに雑音分解後の統計量 $\hat{\mu}_k^{(SH)}$ を用いると，式 (5) より再推定式は，

$$
\begin{aligned}
&\Delta\hat{\mu}^{(H)} \\
&= \frac{\displaystyle\sum_{k=1}^{K}\frac{\sum_p\sum_n\sum_t \gamma_{t,k}^{(p,n)} o_t^{(p,n)} - \gamma_k(\mu_k^{(S)}+\mu^{(H)})}{\sigma_k^{2(S)}+\sigma^{2(H)}}}{\displaystyle\sum_{k=1}^{K}\frac{\gamma_k}{\sigma_k^{2(S)}+\sigma^{2(H)}}} \\
&= \frac{\displaystyle\sum_{k=1}^{K}\gamma_k\frac{\hat{\mu}_k^{(SH)}-\mu_k^{(S)}-\mu^{(H)}}{\sigma_k^{2(S)}+\sigma^{2(H)}}}{\displaystyle\sum_{k=1}^{K}\frac{\gamma_k}{\sigma_k^{2(S)}+\sigma^{2(H)}}}
\end{aligned}
$$

となる．このように，時系列パラメータを用いるのではなく，その統計量 $\hat{\mu}^{(SH)}$ を用いて音響伝達特性のモデルパラメータ $\Delta\hat{\mu}^{(H)}$ を求める．

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，

$$
\sum_k^K \gamma_k \frac{\sigma_k^{2(S)}+\sigma^{2(H)}+\Delta\sigma^{2(H)}-\phi_k}{(\sigma_k^{2(S)}+\sigma^{2(H)}+\Delta\sigma^{2(H)})^2} = 0
$$

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，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，この式を原点におけるテイラー展開を行い，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

### 4.1 実験条件

残響だけの環境下（Noise-free）における，HMM分解法の評価を行うために，図4に示される簡易音響実験室（残響時間 約180 ms）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．更に，実環境評価実験のために，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16.7 dBである．

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．以上の条件のもとで，以下の実験を行う．

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

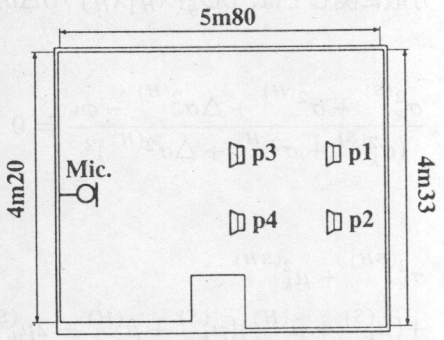


図4 簡易音響実験室
Fig. 4 Experiment room environment.

### 4.2 実験結果

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.)" 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．このとき，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．

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

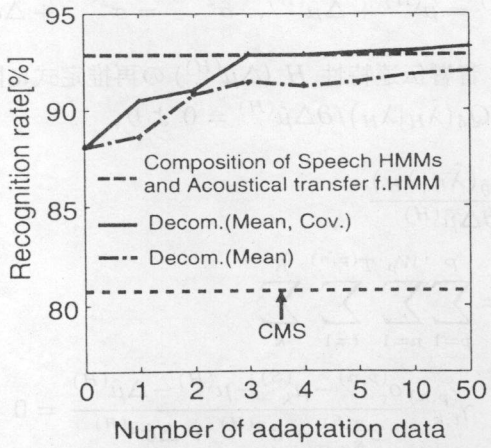


図5 残響環境下（Noise-free）での認識結果
Fig. 5 Recognition rates in reverberant environment.

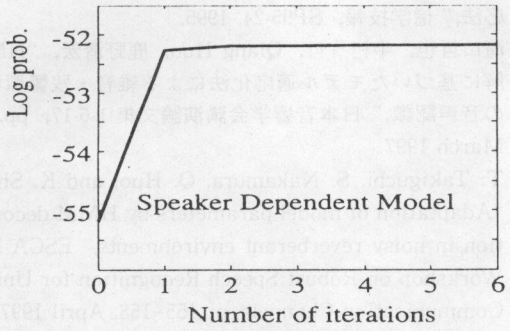


図 6　HMM 分解法の収束性
Fig. 6　Convergence of HMM decomposition.

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づいているのがわかる．また，“CMS”（Cepstral Mean Subtraction）との比較実験では，残響時間が 180 ms の場合，全く効果がないのがわかる．ここでの “CMS” は，1 単語ごとにケプストラム平均値を計算している．

図 6 に平均対数ゆう度とアルゴリズムの反復回数を示す．HMM 分解法では，3 回くらいの繰返しでゆう度が収束しているのがわかる．

次に，実環境下での認識を行う．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，その認識率に近づ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

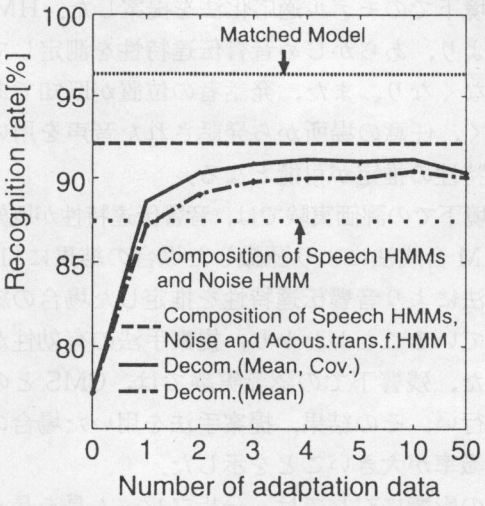


図 7　実環境下での認識結果（特定話者認識）
Fig. 7　Speaker dependent recognition rates in real environment.

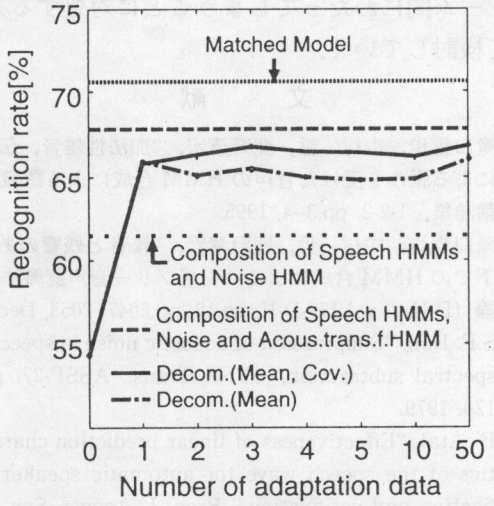


図 8　実環境下での認識結果（不特定話者認識）
Fig. 8　Speaker independent recognition rates in real environment.

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，まだ不十分であることがわかる．これは，フレーム長よりも長いインパルス応答の影響を，分散で完全に吸収することができなかったことによると考えられる．

## 5. むすび

本論文では，HMM 分解・合成法による雑音および

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づいていることがわかり，提案手法の有効性が示せた．また，残響下での認識実験では，CMSとの比較実験を行い，その結果，提案手法を用いた場合の方が更に認識率が大きいことを示した．

残響の影響に対しては，分析フレーム長を長くすることである程度対処できる．しかし，分析窓長を長くすると窓内の定常性の仮定がくずれ，音声のパラメータ抽出の精度が劣化してしまう．今後は，残響の影響がフレーム間にわたってしまうことに対処する方法について検討していく．

## 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] B. Atal, "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, Y. Minami, 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 S.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, "Robust speech recognition in additive and convolutional noise using parallel model combination," Computer Speech and Language, vol.9, pp.289–307, 1995.

[8] A. Sankar and C.-H. Lee, "Robust speech recognition based on stochastic matching," Proc. ICASSP, pp.121–124, 1995.

[9] A. Sankar and C.-H. Lee, "A maximum-likelihood approach to stochastic matching for robust speech recognition," IEEE Trans. Speech and Audio Processing, vol.4, no.3, pp.190–202, 1996.

[10] 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, 1995.

[11] 滝口哲也，中村　哲，Qiang Huo，鹿野清宏，"HMM分解に基づいたモデル適応化法による雑音・残響環境下での音声認識，" 日本音響学会講演論文集 1-6-17, pp.39–40, March 1997.

[12] T. Takiguchi, S. Nakamura, Q. Huo, and K. Shikano, "Adaptation of model parameters by HMM decomposition in noisy reverberant environments," ESCA-NATO Workshop on Robust Speech Recognition for Unknown Communication Channels, pp.155–158, April 1997.

[13] M. Afify, Y. Gong, and J.P. Haton, "A unified maximum likelihood approach to acoustic mismatch compensation: Application to noisy lombard speech recognition," Proc. ICASSP, pp.839–842, April 1997.


（平成9年11月4日受付，10年3月26日再受付）


**滝口　哲也**

1994岡山理大・理・応用数学卒．1996奈良先端科学技術大学院大学情報科学研究科博士前期課程了．現在，同博士後期課程在学中．音声認識の研究に従事．日本音響学会会員．



**中村　哲**　（正員）

昭56京工繊大・工芸・電子卒．昭56～平6シャープ（株）中央研究所および情報技術研究所に勤務．昭61～平1 ATR自動翻訳電話研究所に出向．平6より奈良先端科学技術大学院大学情報科学研究科助教授．平8年3～8月Rutgers University・CAIP Center客員教授．音声情報処理，主として音声認識の研究に従事．京都大学博士（工学）．平4日本音響学会粟屋学術奨励賞受賞．IEEE，情報処理学会，日本音響学会，人工知能学会各会員．



**鹿野　清宏**　（正員）

昭45名大・工・電気卒．昭47同大大学院修士課程了．同年電電公社武蔵野電気通信研究所入所．昭59～61カーネギーメロン大客員研究員．昭61～平2 ATR自動翻訳電話研究所音声情報処理研究室長．平4 NTTヒューマンインタフェース研究所主席研究員．平6より奈良先端科学技術大学院大学情報科学研究科教授．音情報処理学講座を担当．工博．主として音声・音情報処理の研究および研究指導に従事．昭50本会米沢賞．平3 IEEE SP 1990 Senior Award，平6日本音響学会技術開発賞．IEEE，音響学会，情報処理学会各会員．