

統計的音韻中心の推定とそのHMMへの適用

大川 茂樹 白井 克彦

早稲田大学 理工学部 電気工学科

〒169 東京都新宿区大久保3-4-1

E-mail: ookawa@shirai.info.waseda.ac.jp

あらまし 不特定話者・大語彙向けの新たな音声認識手法として、統計的音韻中心の概念と隠れマルコフモデル(HMM)の技術を融合した音韻中心間HMMについて、その理論的背景および評価実験の結果を述べる。まず、全ての音韻に対して統計的に決定され得る中心が存在するとの仮定を設け、局所尤度の最大化に基づく繰り返し学習によりその点を推定する。次に、得られた音韻中心間セグメントに対してHMMを学習し、その連結により単語HMMを作成する。中心尤度を考慮した最適化アルゴリズムにより評価実験を行なった結果、単語認識率の向上を確認した。

キーワード 音韻認識・単語認識・統計的音韻中心・隠れマルコフモデル

Estimation of Statistical Phoneme Center Applied to Hidden Markov Model

Shigeki Okawa Katsuhiko Shirai

Department of Electrical Engineering, Waseda University
3-4-1 Okubo, Shinjuku-ku, Tokyo, 169 Japan

Abstract This paper proposes a new approach for speech recognition which combines a concept of statistical phoneme center and the conventional HMM technique. First, the imaginary center is assumed for each phoneme and it can be estimated by an iterative training mechanism based on the maximization of the local likelihood. Next, the HMMs are trained for every segment between the centers and are concatenated to build the word HMMs. As the experimental result, the word recognition accuracy is improved when the optimal algorithm considering the center likelihood is applied.

Keywords phoneme recognition, word recognition, statistical phoneme center, hidden Markov model

1. はじめに

音声認識において、音響情報に基づく音韻性抽出の技術は、上位レベルの処理に重要な影響を及ぼすために、十分な信頼性を持たせておく必要がある。特に、言語モデルや文法等のトップダウン情報を厳密に利用できない自由な発話に対しては、認識性能はボトムアップ情報に大きく左右される。しかるに近年の音声認識研究では、もはや音韻性の抽出技術の追求は重要視されず、既存の技術がそのまま流用されるにとどまることが多い。

我々はこれまで、特に音声中の音響情報と音韻性との相互情報量に着目した音韻性抽出技術について検討してきた結果、認識性能をある程度まで向上させることに成功した [1, 2]。しかし、理論的に厳密さを欠いた部分が存在し、周囲環境や話者性の大きな変動に十分に対処できないというような問題も依然として残されていた。

最近、統計的な考え方に基づく新たな音韻性抽出手法を考案し、いくつかの予備実験によりその有効性を示した [3]。この方法では、すべての音韻には統計的な中心が存在するとの仮定を設け、前後の音響特徴からその点を推定することにより音韻性を抽出する。今回は、この統計的音韻中心の概念と隠れマルコフモデル (HMM) の技術とを融合することを試みた。以下、従来法との比較を含めてその理論的アプローチを紹介した後、HMM への適用法について述べる。

2. 統計的手法による音韻性抽出

まず、統計的 (確率的) 手法による音韻性の抽出について、若干の基本事項の確認および記号の定義を行う。ここで述べるのは、単語や文章を単位とした最適化の技術は含まず、局所的な一時点における音韻性の抽出に関してであることに注意されたい。

2.1 準備

音響分析により得られた特徴ベクトル系列を \mathbf{X}_t ($t = 0, 1, \dots, T$) とする。時刻 t の音響特徴 \mathbf{X}_t が得られたときに、その時点の音韻カテゴリが Y であるという事象の組をたくさん収集し、適切な方法によりその確率分布を計算しておく、未知の音声入力に対して、条件つき確率 (または条件つき確率密度) $p(Y|\mathbf{X}_t)$ が定義できる。

通常はこの確率の値をサンプルから直接計算することができないため、バイズの定理

$$p(Y|\mathbf{X}_t) = \frac{p(Y)p(\mathbf{X}_t|Y)}{p(\mathbf{X}_t)} \quad (1)$$

の右辺分子の最大化問題に帰着させる。この式で、 $p(Y)$ は、音韻 Y が生起する事前確率であり、言語的な条件に依存する。また $p(\mathbf{X}_t|Y)$ は、音韻 Y が観測されたときの音響特徴 \mathbf{X}_t の生起確率 (事後確率) であり、大量の学習データを観測することで容易に推定できる。分母の $p(\mathbf{X}_t)$ は計算できないが、これは Y に依存しないので分子の最大化には無関係である。

2.2 従来の方法

次に、従来我々が用いてきた音韻性抽出手法について簡単に述べる。

実際の音声では、時刻 t の音響特徴はその前後時刻の音響特徴の影響を十分に受けるため、確率の最大値 $\max p(Y|\mathbf{X}_t)$ を与える音韻 Y をその時点の音韻性抽出結果とするには無理がある。時間的な変化を最も簡単に取り入れるには、直前の時刻の特徴も同時に考える、すなわち確率 $p(Y|\mathbf{X}_{t-1}, \mathbf{X}_t)$ を計算すればよい。

これは、音声に単純マルコフ過程を導入したの他に他ならない。よく知られているように、HMM ではこの確率の計算を状態間の遷移確率と観測シンボルの出力確率とに帰着させて、効率のよい認識アルゴリズムが実現されている。

しかし、実際の音声パターンでは、ある時刻の音韻性を決定付ける要因は、直前の時刻のみならず、かなり前の時刻あるいは未来の時刻の情報にも含まれると考えられる。すなわち、確率

$$p(Y|\dots, \mathbf{X}_{t-2}, \mathbf{X}_{t-1}, \mathbf{X}_t, \mathbf{X}_{t+1}, \mathbf{X}_{t+2}, \dots)$$

を計算するのが本来ならば理想的である。

これを忠実に計算するのは、計算量が膨大になる上、無限個の学習サンプルを必要とするので現実には不可能である。そこで我々は、各時刻の音響特徴は独立に生起する (実際には相関がある) との近似を施し、この確率を $\prod_k p(Y|\mathbf{X}_{t \pm k})$ として扱うものとした。もちろん、入力音声の全区間の確率分布を考慮するわけにはいかないので、適当な位置 (着目点の前後 5~10ms) までで計算を打ち切る。また、確率の乗算を繰り返す行くとアンダーフローが生じるので、実際には対数尤度として計算する。

この方法での問題点は、ある音韻に対する確率分布の学習をその音韻の全ての区間に渡って平等に行うので、過渡的な部分のいわば汚れた特徴をも捨ってしまうことにあった。多くの場合、過渡部の情報はより安定した定常部の情報によって隠蔽されるから、これは常に重大な問題とはならないが、音韻の種類によっては、安定性が小さい場合もある。そこで我々は、ベクトル量子化により音韻カテゴリの条件つきエントロピーを計算し、これを重み係数として用いて、この問題に対処してきた。ここでは、時刻 t が音韻 Y に含まれる尤度が式(2)で計算された:

$$S(t, Y) = \frac{1}{n_f + n_b + 1} \sum_{k=-n_f}^{n_b} \frac{\log p(Y|X_{t+k})}{H(Y|X_{t+k})} \quad (2)$$

$$\left\{ \begin{array}{l} X_{t+k} : \text{時刻 } t+k \text{ の音響特徴ベクトル} \\ H(Y|X_{t+k}) : X_{t+k} \text{ に対する音韻カテゴリの} \\ \quad \text{条件つきエントロピー} \\ n_f, n_b : \text{考慮する前/後フレーム数} \end{array} \right.$$

3. 統計的音韻中心

上述のような問題が生じる原因は、音韻が時間的な区間を持つ単位であるとの固定観念に捉われていたことにあると思われる。いいかえると、ある時刻の音韻性を確率的に表現するときに、その時点がどんな音韻の途中にあるのかを考えることに目が向いていたのである。

このような発想を捨て、利用すべき情報の最適な選択方法を考えた結果、音韻区間中のある時点にその音韻を代表させ、常にその一点を推定する問題に帰着させるやり方が非常に都合よいことが分かった。

これが、以下に述べる統計的音韻中心の考え方の基礎である。このような考え方により、音韻から音韻への渡り部分のような過渡的な区間の情報を低く抑え、逆に非常に安定してその音韻らしさを多く含む部分を十分に重み付けするという機構が、自動的に実現できる。

3.1 定義

まず初めに、統計的音韻中心（以下音韻中心と書く）の定義として、次のような仮説を設ける。

- (1) すべての音韻がその中心点を1つ持つ。
- (2) 音韻中心は、その近辺の音響特徴により一意に推定される。

この仮説に酷似した考え方は、今から数十年前によく提唱されていた。当時の仮説は、あらゆる音韻にはその音韻性が顕著に集中した特異な点（不変的な手がかり）が存在する、というものであり、特に音声知覚の分野では、音韻の切断実験などによりその存在を追及する試みが多くなされた[4]。

ここで注意しなければならないのは、それら過去の考え方と我々の仮説とは、一見類似しているが、根本的に全く異なるということである。ここでの我々の考え方は、決して不変的な音韻性が集中した特別の点を見つけようとするものではなく、あくまでも統計的な方法で定められる一点を推定しようとするに過ぎない。だから、仮にその点が多少前後にずれても、それによって推定精度が大きく変動するようなことはなく、またこの中心点のみであらゆる音韻性を表現しようとしているのでもない。

3.2 音韻中心尤度の計算

音韻中心は、大量の音声データに対して後述の繰り返し学習を行うことで推定される。このとき、中心位置の収束性の評価基準として、音韻中心らしさを示す音韻中心尤度を定義する。この尤度は同時に、未知入力音声に対する認識処理の過程でも使用される。

音韻中心尤度は、前に述べた統計的な考え方に基づき、前後の時刻（分析フレーム）を含めた音響特徴に対する中心位置の事後確率で表現される。式(2)の従来法との違いは、確率計算の対象が音韻 Y （この音韻の区間内であるかどうか）ではなく、中心位置 y^* になることである。

時刻 t に音韻 Y の中心が存在する尤度は次式で計算される:

$$S^*(t, Y) = \frac{1}{n_f + n_b + 1} \sum_{k=-n_f}^{n_b} \log p(y^*|X_{t+k}) \quad (3)$$

$$\left\{ \begin{array}{l} y^* : \text{音韻 } Y \text{ の中心位置} \\ X_{t+k} : \text{時刻 } t+k \text{ の音響特徴ベクトル} \\ n_f, n_b : \text{考慮する前/後フレーム数} \end{array} \right.$$

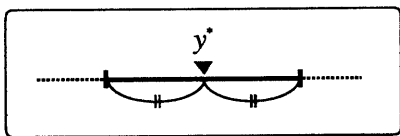
3.3 学習アルゴリズム

次に、音韻中心尤度を計算するための確率分布を学習する方法について述べる。学習データを用いて確率 $p(y^*|X_t)$ の分布を推定するには、真の音韻中心位置を教師信号として与えればよい。しかし、この中心位置は音韻ラベルとは異なり事後的に決定さ

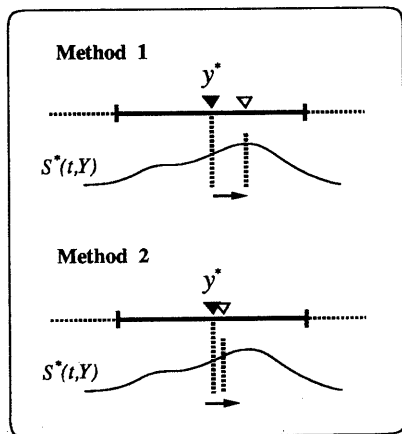
れるものであるから、初期学習データに対しては未知である。

そこで、様々な検討の結果、次のような繰り返し学習のアルゴリズムを考案し、中心尤度の収束性を検証することにより音韻中心を決定する。

- (1) まず、適当量の学習サンプルに対して、何らかの方法で音韻ラベルを与える。視察ラベルが付与されている場合はそれを用いればよいし、適当な自動ラベリングにより付与されたラベルでもよい。このラベルの厳密さは要求しない。与えられた音韻区間の二等分点を初期中心位置とし、マークを付す。



- (2) すべての y^* に対して、その前後数フレーム内の音響特徴 $X_{t \pm i}$ の分布をそれぞれ計算する。確率分布は、離散分布または連続分布（正規分布・混合正規分布）として扱う。
- (3) 得られた確率分布を用いて、もとの中心位置近傍の中心尤度 $S^*(t, Y)$ を各音韻カテゴリに対して計算する。
- (4) 中心尤度の局所最大値（もとの中心位置から ± 10 フレーム（約 100ms）以内の極大値）を与える点を次の音韻中心位置とする。この際、もとの中心位置が正確でない可能性があることを考慮し、次の2通りの方法により推定中心位置を移動する。



方法1：局所最大値をそのまま次の中心位置とする。

方法2：もとの中心位置を局所最大値の方向に少しずつした点（具体的には1フレーム移動した点）を次の中心位置とする。

- (5) 中心位置および中心尤度が収束するまで、(2)~(4)の手続きを繰り返す。

4. 音韻中心推定実験

上の学習アルゴリズムを大量の音声データに適用し、音韻中心の収束性の調査およびその推定実験を行う。

4.1 実験条件

今回行った実験の条件を以下に示す。これらは、以下の全ての実験について共通である。

}	音声試料	ATR 日本語音声データベース 日本語重要語 5240 単語 音韻バランス 216 単語
	話者	男性 10 名
	A/D 変換	12.0kHz 標本化, 16bit 量子化
	高域強調	1 次差分フィルタ
	フレーム長	21.3ms, ハミング窓
	分析周期	5.0ms
	考慮フレーム	$n_f = n_b = 3$ (予備実験より)
	音響特徴量	LPC メルケプストラム [1-16]
	音韻カテゴリ	無音を含む 27 種

音韻中心推定実験は、各話者ごとに独立に行なった（特定話者実験）。学習には日本語重要語 5240 単語のうち奇数番目の単語を、評価には偶数番目の単語を用いた。離散確率分布を用いる場合のベクトル量子化用コードブックは、音韻バランス 216 単語により作成した。

式 (3) の確率分布 $p(y^* | X_t)$ としては、次の4種類を用いてその結果を比較する。

}	D1	離散分布 (量子点数 256, 通常の VQ による)
	D2	離散分布 (量子点数 2528, 相互情報量を考慮した階層的クラスタリング [5] による)
	C1	連続分布 (無相関単一ガウス分布)
	C2	連続分布 (無相関混合ガウス分布, 混合数 4)

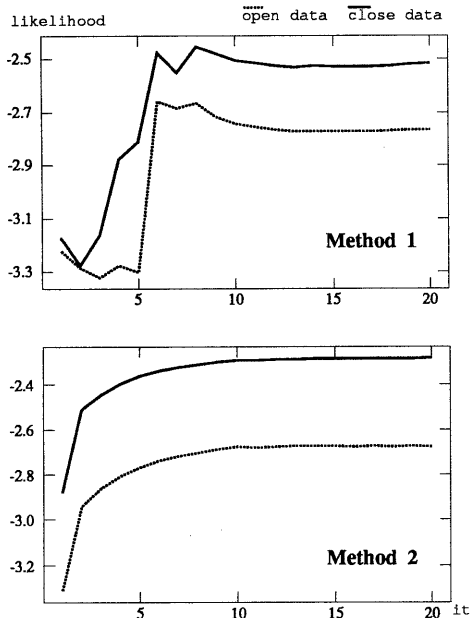


図1: 中心尤度の収束性

4.2 中心尤度の収束性

まず、繰り返し学習により中心尤度がどのように推移するかを調べた。提案した2通りの方法により音韻中心を移動しながら、学習を20回繰り返したときの中心尤度の変化を図1に示す(話者 MAU, D2, 全音韻平均値)。この際、初期中心位置により学習した場合の尤度を1回目と数えている。また、同じ

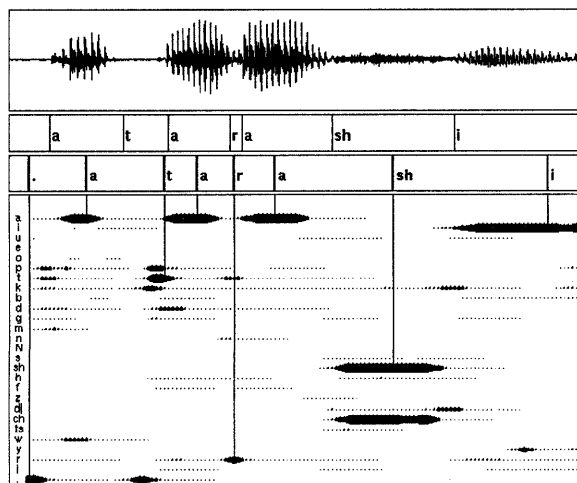


図2: 音韻中心尤度の例

データセットに対して学習を繰り返すと、そのデータに依存した学習をしてしまう恐れがあり、厳密な評価ができないので、異なるデータ(図中で open data)に対する尤度も同時に計算している。

図1を見ると、局所最大値に直接移動する方法1よりも、少しずつ移動させる方法2の方が収束が滑らかで早いことが確認できる。また、異なるデータセットに対する尤度も同様に上昇かつ収束していることから、提案する学習法の有効性が確認できる。

図2には、推定された音韻中心および中心尤度の例を示す(話者 MAU, D2, 単語 /atarashii/, 学習10回)。

4.3 中心検出性能

次に、音韻中心尤度の極大値に対して適当な閾値を設定して簡単な中心検出実験を行う。中心尤度が閾値を超えたときにその音韻が存在するとして正解音韻ラベルと比較する。閾値を下げると、次第に検出される音韻中心は増えるが、他の対立候補音韻もまた検出されてしまうので挿入誤り数が増える。理想的には、高い中心検出率に対して低い挿入誤り率が達成できればよい。実験結果を図3に示す(各話者の平均値、繰り返し学習10回)。

図3より、D2およびC2の場合に他の分布に比べてよい結果が出ていることが分かるが、この2つの間での有意差は見られない。性能としては、90%の検出率に対して32%の誤り率が、50%の検出率に対して15%の誤り率が達成できていることが確認できる。

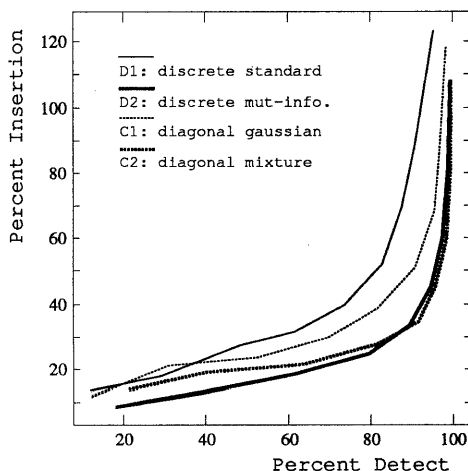


図3: 中心検出率と挿入誤り率の関係

5. 音韻中心間 HMM

以上に、統計的音韻中心の概念を用いた音韻性抽出に関する理論と実験結果を示した。この手法を実際の音声認識に適用するにはいくつかの方法が考えられる。定義の項で述べたように、音韻中心は音韻性抽出のための時間的な一点であり、実際に認識したいのは時間的な長さを持つ単語や文章であるから、時区間を記述できるモデルを導入するのが望ましい。この最も容易な方法は、音韻中心尤度をそのまま距離尺度として用いて、DP による時間軸正規化を行うものであり、この実験結果は既に報告している [3]。

今回は、ある音韻中心と次の音韻中心の間に隠れマルコフモデル (HMM) を適用して単語単位での最適化を行うことを考える。HMM は、その統計処理のしやすさと不特定話者などの環境変動への頑健性から、近年の音声認識システムによく用いられている。特に、大量の音声データを用いた連結学習の機構が開発された結果、効率よくモデルを作成できるようになり、よい性能を与えている。

音韻中心間のセグメントを単位として HMM を構成するという考え方は、渡辺らによって提案されている半音節 HMM とある意味で類似している [6]。この研究では、音節を母音の中心で二つに分割したもの、として「半音節」を定義しており、半音節セグメントの境界としては、例えば無声破裂音の場合は破裂開始点、半母音や流音の場合は子音中央の振幅極小点というように基準を定めている。この境界点は、我々の考えている音韻中心に結果的にはほぼ相当すると思われる。しかし、音韻中心はあらかじめ明示的に定義することができない上、必ずしも中心間に挟まれるセグメントが音節の半分に相当するわけではないので、半音節との用語を当てるのは適当でない。そこで本研究では、音韻中心間のセグメントを単位とした HMM を音韻中心間 HMM と呼ぶことにする。

5.1 音韻中心間 HMM の学習

音韻中心間 HMM は、音韻中心の推定アルゴリズムにより決定された中心位置により学習する。2つの音韻の組み合わせで1つのモデルが構成されるので、モデル数は標準的な音韻を単位としたモデルよりも多くなるが、日本語が基本的に CV (子音+母音) 音節から構成されることを考えると、音韻数の自乗よりはるかに少なくてすむ。今回は、音韻カ

テゴリとして、{a, i, u, e, o, p, t, k, b, d, g, s, sh, h, f, z, dj, ch, ts, m, n, N, w, y, j, r, .} の 27 種類を用いた (. は無音を表す)。表 1 にそれらの組み合わせにより作成された音韻中心間 HMM の種類を接続構造別に示す。ここで、V は母音、C は子音、S は無音を表す。この分類は、CV 音節の他に長母音セグメントや k-s, f. のような無声化を考慮した音韻間セグメントも含んでいる。

表 1: 音韻中心間の種類

構造	個数
V-V	28
V-C	73
C-V	75
C-C	10
V-S	5
C-S	8
S-V	5
S-C	19
S-S	1
合計	224

5.2 単語モデルの作成

学習された音韻中心間 HMM を連結して単語 HMM を作成する。連結は、できる限り発声変形 (無声化など) を考慮した音韻連鎖記述に基づいて行うが、厳密な音韻環境の記述はしていない。

図 4 に、音韻中心間 HMM の連結の様子を示す。 $M(i, j)$ は、音韻中心 y_i^* と y_j^* の間のセグメントにより学習した HMM を表す。各 $M(i, j)$ の連結部において、音韻中心 y_j^* を通過する構造となる。連結は、前の HMM の最終状態を次の HMM の初期状態に重畳して実現している。そのため、音韻中心間 HMM の最終状態は自己遷移を持たない。

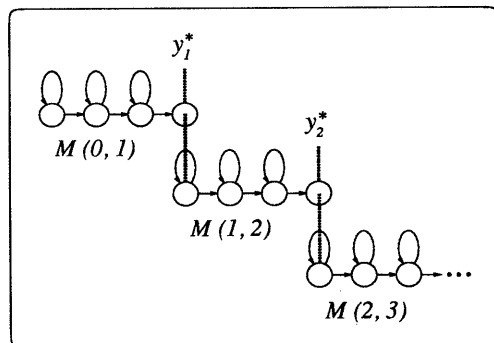


図 4: 音韻中心間 HMM の連結

5.3 中心尤度を考慮した評価アルゴリズム

さて、音韻中心間 HMM の連結モデルを用いて単語認識を行うにあたり、先に定義した音韻中心尤度を積極的に利用することを考える。

音韻中心尤度と HMM の尤度とは当然独立ではなく、非常に複雑な関係にあると思われるので、これらを確率的に組み合わせることは難しい。この問題の最も簡単な解決策として、本研究では、モデルの評価時に、音韻中心尤度を適当に重み付けた上で HMM の局所累積尤度に加算し、単語全体の最適化を行う。どのような重み係数を用いればよいか分からないので、適当に変化させて性能を比較する。

以下に、HMM の評価アルゴリズムを示す。今回は Viterbi アルゴリズムを用いているが、Baum-Welch アルゴリズムに対しても同様な処理が可能である。図 5 には、HMM のトレリス上における音韻中心尤度加算の様子を示す。

記号

- a_{ij} : 状態 i から状態 j への遷移確率 ($i, j = 1, \dots, n$)
- $b_i(x_t)$: 状態 i への遷移時にシンボル x_t を出力する確率 ($t = 0, \dots, T$)
- $f(i, t)$: 状態 i , 時刻 t までの累積対数尤度
- c_{ij} : 状態 j が音韻中心間 HMM の連結部であり、 $i = j - 1$ のとき 1; 他のは 0
- $S^*(t, Y_i)$: 連結部の状態 i が対応する音韻の中心尤度 (式 (3) 参照)
- w_c : 音韻中心尤度に対する重み係数

初期化

$$\begin{cases} f(1, 0) = 1 \\ f(i, 0) = 0 \quad (i \neq 1) \end{cases}$$

手続き

for $t = 1, 2, \dots, T$
 for $i = 1, 2, \dots, n$

$$f(i, t) = \max_j [c_{ji} \cdot w_c \cdot S^*(t, Y_i) + f(j, t - 1) + \log a_{ji} b_i(x_t)]$$

最終時刻 T における $f(n, T)$ がこのモデルに対する Viterbi 尤度を与える。

5.4 システム構成

本研究で用いたシステムの構成を図 6 に示す。

学習系では、まず、音響特徴から局所尤度の最大化に基づく繰り返しアルゴリズムにより音韻中心を推定し、収束後の中心位置の確率分布を音韻中心辞書 (PC Dictionary) に格納する。その後、決定された中心間のセグメントに対して、音韻中心間 HMM (PC-HMM) を作成し、それらを連結して単語モデルとする。

認識系では、音韻中心辞書をもとに各フレームの音韻中心尤度を全音韻に対して計算する。その後、前述の評価アルゴリズムに基づき、単語 HMM の尤度を計算し、認識結果とする。

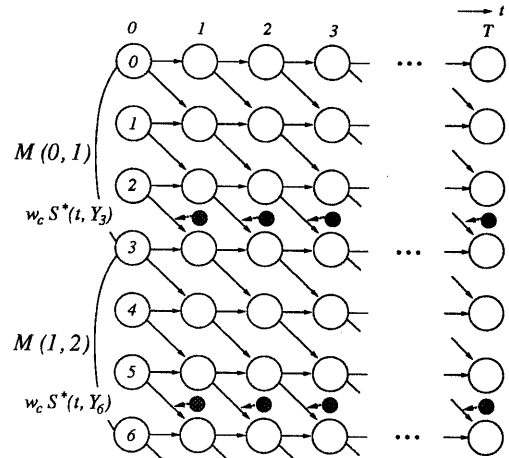


図 5: トレリス上での音韻中心尤度の加算

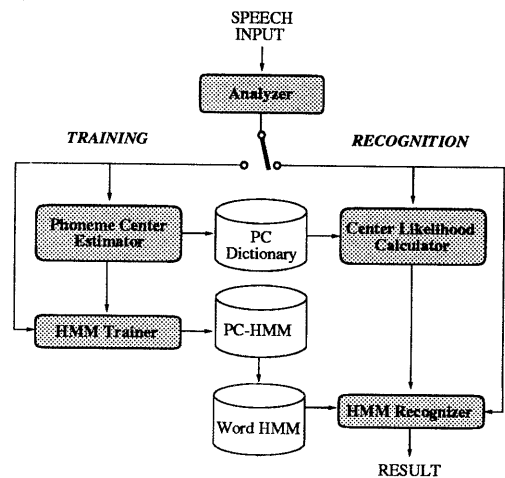


図 6: システム構成

5.5 単語認識実験

音韻中心間 HMM の学習は、10 名話者のうち 9 名の日本語重要語 5240 単語から切り出した音韻中心間セグメントに対して、Baum-Welch アルゴリズムを適用して行なった。HMM の出力確率分布としては、無相関正規分布 (CS) と、無相関混合正規分布 (混合数 4) (CM) の 2 種類を用いて比較した。HMM の状態数は、すべての音韻中心間に対して 3 状態 4 出力分布 (最終状態は自己遷移なし) とした。

この音韻中心間 HMM を連結し、ATR 音韻バランス 216 単語についてモデルを作成する。評価は、学習に用いた話者とは別の 1 話者について行なった。この際に用いる音韻中心辞書は、HMM の学習に用いたのと同じ 9 名話者の 5240 単語についてあらかじめ作成しておいた。

比較のために、同じ条件下で標準的な音韻 HMM (27 カテゴリー) を学習し、同様の実験を行なった。このときの状態数は、4 状態 5 出力分布とした。

表 2 に実験結果を示す。表中で PH は標準的な音韻 HMM を用いたものを、PC は音韻中心間 HMM を用いたものを表す。また、 w_c は HMM の評価アルゴリズムにおける音韻中心尤度に対する重み係数を示す。 $w_c = 0$ のときは、中心尤度を用いていないことを意味する。

表 2 より、標準的な音韻 HMM を用いた場合 (PH) の性能に比べ、音韻中心間 HMM を用いた場合 (PC) に性能が向上していることが確認できる。特に最も認識率の良い $w_c = 5$ の場合、CS で約 8%、CM で約 3% の性能向上が見られた。

表 2: 単語認識実験結果

HMM	w_c	単語認識率 %	
		CS	CM
PH	-	88.4	94.4
	0	90.7	95.4
PC	1	92.6	96.8
	5	96.8	97.7
	10	94.9	96.3
	15	93.5	95.8
	20	91.2	94.0

6. まとめ

本論文では、統計的音韻中心の概念に基づく音韻性抽出手法とその HMM への適用法について、理論的アプローチおよび評価実験結果を報告した。統

計的に推定された音韻中心間セグメントに対して HMM を作成し、それらの連結により単語認識を行なった結果、標準的な方法に比べて単語認識率が 3 ~ 8% 向上し、本手法の有効性が確認された。

今後は、次の各点について検討を進めていく予定である。

- (1) Δ ケプストラム等の時間変化を考慮したパラメータの利用。
- (2) 対立候補との相互間における識別精度向上のためのネガティブ学習方式の開発。
- (3) 音韻中心の音声知覚的、韻律特徴的な分析。
- (4) 音韻論や弁別素性論に基づく音韻中心間カテゴリの最適化。

謝辞 日頃御討論いただく音声言語研究室の諸氏に感謝します。実験に用いた HMM の開発にあたっては、日本情報処理開発協会製作の連続音声データベース利用支援ツールを参考にさせて戴きました。記して謝意を表します。本研究の一部は文部省科学研究費重点領域研究「音声対話」の補助による。

参考文献

- [1] S. Okawa, T. Endo, T. Kobayashi and K. Shirai: "Phrase Recognition in Conversational Speech Using Prosodic and Phonemic Information," *IEICE Trans. Information and Systems*, E76-D, 1, pp.44-50, (1993-1)
- [2] 大川, 小林, 白井: "情報量を基準とした音韻性抽出手法を用いた連続音声中の音韻認識," *音響誌*, 50, 9, pp.425-433, (1994-9)
- [3] 大川, 白井: "音韻性の中心の推定とその有効性の検討," *信学技報*, SP94-17, pp.21-28, (1994-6)
- [4] K. N. Stevens, S. E. Blumstein: "Invariant Cues for Place of Articulation in Stop Consonants," *JASA*, 64, (1978-11)
- [5] 白井, 青木: "音響特徴量の情報量に基づく階層的クラスタリングによる音韻認識," *信学論*, J72-D-II, 8, (1989-8)
- [6] 渡辺, 吉田, 古賀: "半音節を単位とした HMM を用いた大語い音声認識," *信学論*, J72-D-II, 8, (1989-8)