音声言語情報処理 29－6
(1999. 12. 20)

状態遷移に順序関係を持つ非同期遷移型 HMM

松田 繁樹　中井 満　下平 博　嵯峨山 茂樹

北陸先端科学技術大学院大学 情報科学研究科
〒923-1292 石川県能美郡辰口町旭台 1-1
URL http://www-ks.jaist.ac.jp/index-j.html

あらまし　本報告では、個々の特徴間の状態遷移に順序の制約を持つ非同期遷移型 HMM(AT-HMM) を提案する。私達は、個々の特徴が非同期に変化している特徴ベクトルの系列をスカラー出力型 HMM により表現する順序制約無し AT-HMM を以前に提案した。本報告では、更に音声信号を効果的に表現するため、個々の特徴量の状態遷移に順序関係の制約を付けた、順序制約付き AT-HMM について検討を行う。また、順序制約付き AT-HMM の実現法として、"時間方向共有" と呼ぶ新しいパラメータ共有の概念を説明する。切り出し音素認識と孤立単語認識において、従来型 HMM と比較して約 20%と 40%の誤り削減率が得られた。AT-HMM のモデル化能力を更に改善するため、特徴毎に音素環境クラスタリングを行う特徴量別音素環境クラスタリング (FW-PEC) の概念を説明し、FW-PEC を用いた AT-HMM 生成法として、特徴量別逐次状態分割法 (FW-SSS) を提案する。切り出し音素認識と孤立単語認識において、FW-PEC を用いていない AT-HMM よりも更に 10 ポイント程度の誤りが削減された。

キーワード　音声認識, 非同期遷移型 HMM, 時間方向共有構造, 特徴量別音素環境クラスタリング

Asynchronous Transition HMM with Sequential Constraints

Shigeki MATSUDA　Mitsuru NAKAI
Hiroshi SHIMODAIRA　Shigeki SAGAYAMA

School of Information Science
Japan Advanced Institute of Science and Technology, HOKURIKU
1-1 Asahidai, Tatsunokuchi, Ishikawa 923-1292
URL http://www-ks.jaist.ac.jp/

Abstract　We propose Asynchronous-Transition HMM (AT-HMM) with sequentially constrained state transitions among individual features. Previously, we proposed non-sequential AT-HMM that represents with a scalar-output HMM the sequence of feature vectors whose components change asynchronously with each other. Sequential AT-HMMs inctroduces sequential constraints among asynchronized transitions to better represent the asynchrony among features. A new technique of "state tying along time" is introduced to realize the sequential AT-HMM structure. Sequential AT-HMMs gave approximately 20% and 40% lower error rates compared with conventional HMMs in phoneme and isolated word recognition experiments, respectively. To furtther improve the modeling ability of the sequential AT-HMM, we also introduce the Feature-Wise Successive State Splitting (FW-SSS) algorithm based on a concept of Feature-Wise Phoneme Environment Clustering (FW-PEC). The error rates were further reduced approximately by 10 points from those by the above sequential AT-HMM without FW-PEC.

key words　Speech Recognition, Asynchronous Transition HMM, State Tying along Time, Feature-Wise PEC

1 まえがき

我々は，特徴ベクトルの各特徴量が非同期に変化する信号を効果的にモデル化するため，非同期遷移型HMM(Asynchronous Transition HMM: AT-HMM)[1] [2] の枠組を提案した．この枠組の中で，個々の特徴量における状態遷移の順序に一切の制約の無い，順序制約無し AT-HMM の検討を行った．順序制約無し AT-HMM は，個々の特徴量をスカラー出力型 HMM を用いてモデル化し，それぞれの HMM の尤度を独立に計算することにより，各特徴量の自由な状態遷移を実現した．本報告では，更に音声信号を効果的に表現するため，個々の特徴量の状態遷移に順序関係の制約を付けた，順序制約付き AT-HMM について検討を行う．

順序制約付き AT-HMM は，“時間方向共有” と呼ぶ新しいパラメータ共有法により，従来型 HMM の枠組で実現することができる．“時間方向共有” は，音素環境クラスタリング [3][4]，状態の共有 [5][4]，混合成分の共有 [6]，分布パラメータの共有 [7] とは異なる新しいパラメータ共有の概念である．本報告では，“時間方向共有” の概念により順序制約付き AT-HMM を生成し，音声認識性能の評価を行う．更に，AT-HMM のモデル化能力を改善するため，個々の特徴量毎に音素環境クラスタリング [3] を行う，特徴量別音素環境クラスタリング (Feature-Wise Phoneme Environment Clustering: FW-PEC) の概念について説明する．FW-PEC の概念を用いた，AT-HMM 生成法として，特徴量別逐次状態分割法 (Feature-Wise Successive State Splitting: FW-SSS) を提案し，音声認識性能の評価を行う．

第 2 章では，“時間方向共有” による順序制約付き AT-HMM の実現法及び，音素環境依存モデルの生成法を説明し，切り出し音素認識と孤立単語認識実験により音声認識性能の評価を行う．第 3 章では，特徴量毎に音素環境クラスタリングを行う FW-SSS 法を用いた順序制約付き AT-HMM の生成法を説明し，音声認識性能の評価を行う．第 4 章はむすびである．

2 非同期遷移型 HMM

音声認識に一般に用いられている特徴ベクトル (ケプストラム係数等) の時系列信号を観察すると，各特徴量の時間変化は必ずしも同じタイミングで起こっている訳ではない．個々の特徴量の時間変化が非同期な例として，図 1 に音素/k/、先行音素と後続音素が/a/における音声サンプル 10 個の，第 1MFCC と第 2MFCC の時間変化を示す．現在の音声認識システムは，音響分析によって音声波形を特徴ベクトルの時系列信号変換し，特徴空間内の点の軌跡として扱われている．HMM は，特徴空間内の点の軌跡をベクトルを出力とする定常信号源の連鎖としてモデル化している．図 2-(a) は，従来型 HMM を用いて 2 次元のベクトル時系列信号を 4 つの 2 次元ベクトルを出力とする定常信号源でモデル化した例である．従来型 HMM は，ベクトルの各特徴量が全て同じタイミングで状態遷移 (変化) することを仮定したモデルと見ることができる．しかし，個々の特徴量間の時間変化が非同期な信号は，特徴量毎に別々に状態遷移し，時間変化が同期している信号に対しては同時に状態遷移する方がより効果的に特徴ベクトルの時系列信号をモデル化できると考えられる．

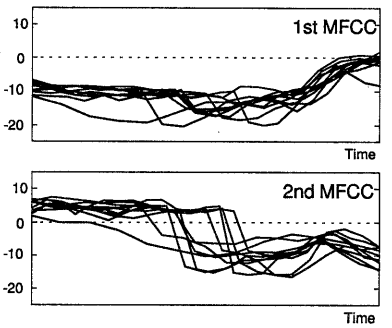

図 1: 音素/k/、先行音素と後続音素が/a/における音声サンプル 10 個の第 1MFCC と第 2MFCC の時間変化

我々は，図 2-(b) のように特徴ベクトルの各特徴量をスカラー出力型 HMM によりモデル化し，個々の特徴量の尤度を独立に計算することにより，特徴間の非同期な状態遷移を実現する順序制約無し AT-HMM を提案した．順序制約無し AT-HMM は，個々の特徴量の状態遷移は独立に計算されるため，特徴間の状態遷移の順序関係は考慮されていない．しかし，図 1 の例において特徴ベクトルの時系列信号は，第 2MFCC で状態遷移が発生し，続いて第 1MFCC で状態遷移するといったように，特徴間の状態遷移に順序関係が存在していると考えるられる．本報告では，各特徴間の状態遷移に順序の関係を付加した順序制約付き AT-HMM について検討を行う．

本章では，個々の特徴量の状態遷移に対して順序の制約を付加した順序制約付き AT-HMM の実現法とモデル生成法の説明を行い，切り出し音素認識及び，孤立単語認識実験により音声認識性能の評価を行う．

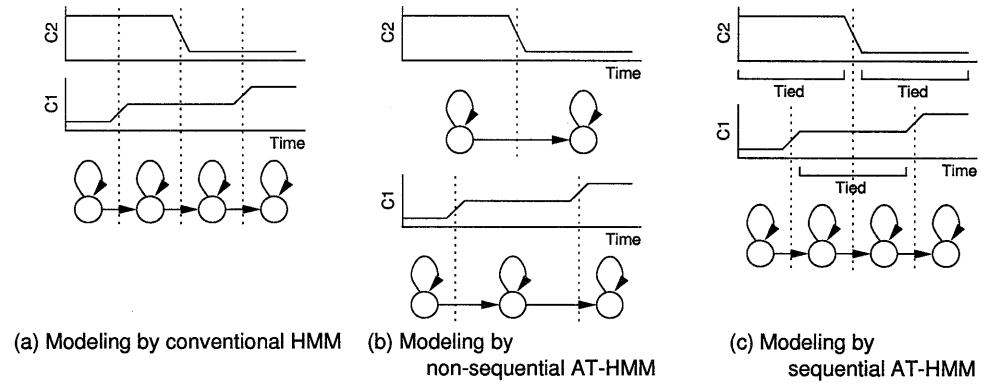

図 2: 従来型 HMM，順序制約無し AT-HMM，順序制約付きによるモデル化

2.1 順序制約付き AT-HMM の実現法

個々の特徴量の状態遷移に対して，順序制約を付けた順序制約付き AT-HMM の実現法を説明する．順序制約付き AT-HMM は，図 2 の信号の場合，特徴 1，特徴 2，特徴 1 の順番で個々特徴量の状態が遷移するモデルである．この順序制約付き AT-HMM は，“時間方向共有” と呼ぶ新しいパラメータ共有法により従来型 HMM の枠組で実現することができる．図 2-(c) に，“時間方向共有” により実現した順序制約付き AT-HMM によるモデル化の例を示す．図のように，特徴 1 は第 2 状態と第 3 状態の分布パラメータが共有され，特徴 2 は第 1 状態と第 2 状態，また第 3 状態と第 4 状態の分布パラメータが共有されている．このように分布パラメータを共有化することにより，個々の特徴の分布の変化に順序関係を付けた状態遷移が実現される．また，従来型 HMM のモデル当りの状態数は，多ければ多い程，特徴ベクトルの時系列信号の時間分解能は増すことになる．

2.2 順序制約付き AT-HMM の生成法

順序制約付き AT-HMM を生成する手法には，以下に示すように 2 通りの方針が考えられる．本章では，手法 2 の，スカラー出力型 HMM を用いた順序制約付き AT-HMM の生成法を説明する．

手法 1: 複数状態の left-to-right 型 HMM を学習し，隣合う状態において分布パラメータを特徴量毎に共有化する．

手法 2: 各特徴量をスカラー出力型 HMM を用いて学習し，それぞれのスカラー出力型 HMM の遷移タイミングをクラスタリングする．

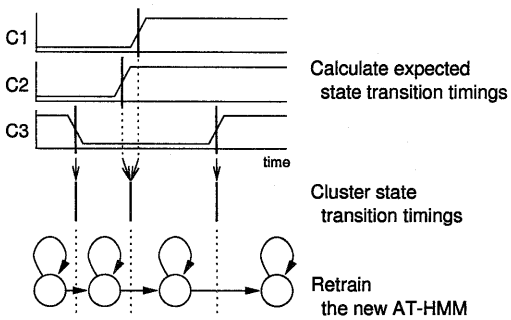

図 3: 順序制約付き AT-HMM のモデル生成法の処理の流れ

モデル当り N 個の状態を持つ順序制約付き AT-HMM の生成法の手順を次に示す．また，図 3 にこの処理の流れを示す．

(1) 従来型 HMM 用に音素環境依存モデルを生成する．

(2) 従来型 HMM の個々の特徴毎にモデルを (スカラー出力型 HMM として) 再学習し，それぞれの特徴量における平均的な状態遷移のタイミングを求める．学習データ中に存在した音素環境全てについて状態遷移確率を再学習を行う．

(3) 学習によって得られた状態遷移のタイミングを $N-1$ 個の代表点にクラスタリングし，個々の特徴量の状態遷移と同じタイミングで状態が変化するように，従来型 HMM の特徴パラメータを時間方向に共有化する．

(5) 得られた時間方向状態共有構造を再学習し，音素環境依存 AT-HMM を生成する．

図 3 の例では，ステップ (3) において，特徴 1 と特徴 2 の遷移タイミングが一つの代表点にクラスタリングされている．また時間方向の共有構造として，特徴 1 と特徴 2 において第 1 状態と第 2 状態，また第 3 状態と第 4 状態が共有されている．同様に，特徴 3 において第 2 状態と第 3 状態の分布パラメータが共有されている．

2.3 音素認識実験

順序制約付き AT-HMM における音素環境依存モデルの音声認識性能を評価するため，特定話者切り出し音素認識実験を行った．AT-HMM は，ML-SSS 法 [8] により生成した従来型 HMM の音素環境依存モデルから生成し，モデル当りの状態数 3, 4, 5, 6, 7, 8 で実験を行った．比較実験として，ML-SSS 法により生成した従来型 HMM と，全ての特徴量の状態共有構造が同じ順序制約無し AT-HMM で認識を行った．ML-SSS 法により生成した従来型 HMM は，学習データに存在した全ての音素環境について，状態遷移確率を付加し再学習を行ったモデルを使用した．モデルは，10600(従来型 HMM で 200 状態)，21200(400 状態)，31800(600 状態)，42400(800 状態) で分布パラメータ数を揃えて実験を行った．尤度計算には Viterbi アルゴリズムを用いた．

学習データには，ATR 研究用日本語音声データベース A-set 中，男性 2 話者 (mht, mau) 女性 2 話者 (fms, ffs) の重要語 5240 単語中の奇数番目と音素バランス単語 216 単語を使用し，評価データには重要語 5240 単語中の偶数番目を使用した．音素ラベルは，/N, a, b, tʃ, d, e, f, g, h, i, ʒ, k, m, n, o, p, Q, r, s, ʃ, t, ts, u, w, j, z/の計 26 音素を用いた．サンプリング周波数 12kHz の波形データをハミング窓 25ms，フレーム周期 8ms で分析した．特徴パラメータは，対数パワー，12 次メル尺度変換ケプストラム，Δ 対数パワー，Δ12 次メル尺度変換ケプストラムの計 26 次元を使用した．

各分布パラメータ数に対する誤り認識率を図 4 に示す．順序制約無し AT-HMM では従来型 HMM よりも認識率が低下しているが，順序制約を付けた AT-HMM では約 20%程度の誤り削減率が得られており，順序制約を付加することにより音素サンプルの識別能力が向上したと考えられる．また，全モデルパラメータ数 (状態遷移確率と分布パラメータ) がほぼ同数の，モデル当り 3 状態の HMM でも同様に 20%近い誤り削減率が得られた．

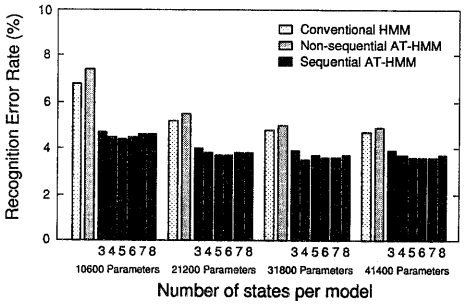

図 4: 順序制約付き AT-HMM，順序制約無し AT-HMM，従来型 HMM における音素認識実験の結果

2.4 孤立単語認識実験

順序制約付き AT-HMM の音素環境依存モデルによる孤立単語認識実験を行った．比較として従来型 HMM の認識実験を行った．順序制約付き AT-HMM の環境依存音素モデルは，切り出し音素認識実験で生成したモデルで最も高い認識率の得られた，モデル当り 5 状態の順序制約付き AT-HMM を使用した．単語辞書としては，重要語 5240 単語中偶数番目を使用し，評価には同じ偶数番目の孤立単語データを使用した．単語尤度は，Viterbi アルゴリズムにより計算した．分布パラメータ数は，10600，21200 で認識実験を行った．

表 1: 従来型 HMM と比較した順序制約付き AT-HMM の誤り認識率と誤り削減率

Method	#parameters	%errors	%reduction
HMM	10600	8.1	—
AT-HMM	10600	4.3	46.9
HMM	21200	6.2	—
AT-HMM	21200	3.8	38.7

各分布パラメータ数に対する誤り認識率と誤り削減率を表 1 に示す．順序制約付き AT-HMM は，従来型 HMM と比較して約 40%程度の誤り削減率が得られた．

3 特徴量別音素環境クラスタリング

音声認識に用いられる特徴パラメータは，ケプストラムやパワー，またその時間変化成分などが用いられている．これらの異なるパラメータのベ

クトル時系列信号は，個々独立な時間変化をしており，それぞれの特徴量をモデル化するために適切な状態数や，状態共有構造は異なっている可能性がある．本章では，特徴ベクトルの時系列信号の個々の特徴毎に，音素環境クラスタリング [3] を行う特徴量別音素環境クラスタリング法 (FW-PEC) を用いた順序制約付き AT-HMM の生成法を説明し，音声認識性能の評価を行う．

3.1 特徴量別逐次状態分割法

FW-PEC の概念を用いた AT-HMM の生成法として特徴量別逐次状態分割法 (FW-SSS) の説明を行う．特徴ベクトルの個々の特徴量に対して最適な状態数や状態共有構造を解析的に求めることは一般に困難である．FW-SSS は，個々の特徴量をスカラー出力型 HMM によりモデル化し，全ての状態に対して逐次状態分割法 (SSS) を基礎とした手法により，時間方向分割と環境方向分割を繰り返しながら除々にモデルを精密化して行く手法である．FW-SSS 法の処理の流れを図 5，手順を次に示す．

(1) 初期 HMnet を作成し全ての学習データを用いて学習する．

(2) 全ての状態に対して，音素環境と時間方向の両方の分割ゲインを計算し，分割によって最大の分割ゲインが得られる状態を見付ける．

(3) 状態の分割を行い，分割によって影響を受けた状態を再学習する．

(4) 必要な状態数まで分割が進むまでステップ 2 と 3 を繰り返し行う．

上記の FW-SSS 法により，個々の特徴量に対応するスカラー出力型 HMM は，それぞれの特徴をモデル化するのに適当な状態数と状態共有構造を持つと考えられる．得られたスカラー出力型 HMM を用いて，第 2.1 章で示した順序制約付き AT-HMM の生成法を用いることにより，個々の特徴量毎に状態数や状態共有構造の異なる順序制約付き AT-HMM が得られる．

3.2 音素認識実験

FW-SSS 法を用いて生成した順序制約付き AT-HMM の，特定話者切り出し音素認識実験を行った．比較のため，ML-SSS 法により生成した従来型 HMM と，順序制約無し AT-HMM の認識実験も行った．順序制約無し AT-HMM は，FW-SSS

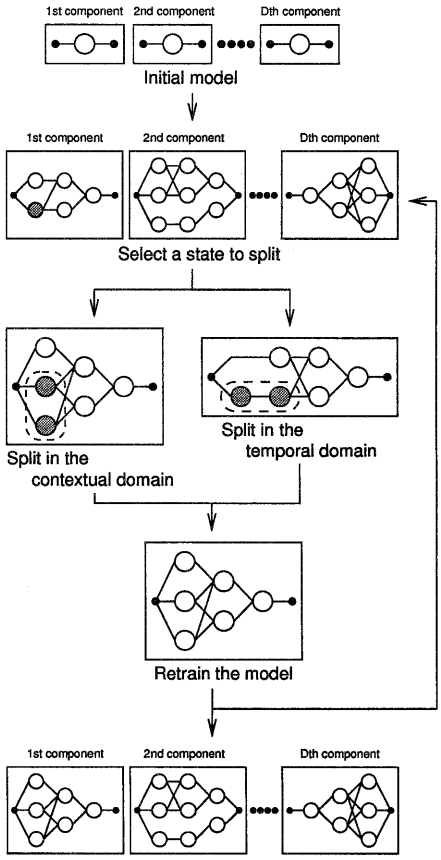

図 5: FW-SSS 法の処理の流れ

法により生成したスカラー出力型 HMM を使用した．実験データは第 2.2 章で行った音素認識実験と同じ条件で分析した．

順序制約付き AT-HMM と順序制約無し AT-HMM，従来型 HMM の誤り認識率を図 6 に示す．順序制約付き AT-HMM の場合は，従来型 HMM と比較して約 30% 程度の誤り削減率が得られた．特徴量別逐次状態分割法により生成された順序制約付き AT-HMM については，各特徴量の状態共有構造が同じ AT-HMM よりも更に 10 ポイント程度高い誤り削減率が得られた．

3.3 孤立単語認識実験

順序制約付き AT-HMM の音素環境依存モデルによる孤立単語認識実験を行った．比較として従来

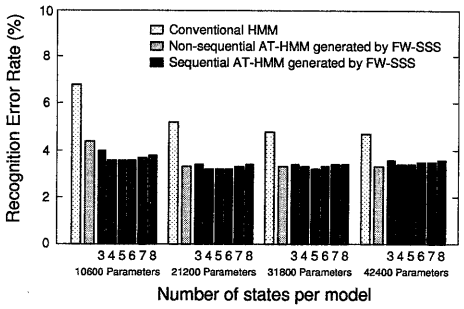

図 6: FW-SSS 法を用いて生成した順序制約付き AT-HMM と順序制約無し AT-HMM，従来型 HMM における音素認識実験の結果

型 HMM の認識実験を行った．環境依存音素モデルは，切り出し音素認識実験で生成したモデルで最も高い認識率の得られた，モデル当り 5 状態の順序制約付き AT-HMM を使用した．単語辞書としては，重要語 5240 単語中偶数番目を使用し，評価には同じ偶数番目の音声データを使用した．単語尤度は，Viterbi アルゴリズムにより計算した．

各分布パラメータ数に対する誤り認識率と誤り削減率を表 2 に示す．特徴量別逐次状態分割法により生成した順序制約付き AT-HMM は，従来型 HMM と比較して 50% 以上の誤り削減率が得られた．

表 2: 従来型 HMM と比較した順序制約付き AT-HMM の誤り認識率と誤り削減率

Method	#parameters	%errors	%reduction
HMM	200	8.1	—
AT-HMM	200	3.2	60.5
HMM	400	6.2	—
AT-HMM	400	3.0	51.6

4 むすび

本報告では，個々の特徴量の状態遷移に対して順序の制約を付けた順序制約付き AT-HMM について議論した．音素環境依存の順序制約付き AT-HMM を生成する手法として，“時間方向共有” と呼ぶ新しいパラメータ共有法を説明した．順序制約付き AT-HMM の音声認識性能を評価するため，切り出し音素認識と孤立単語認識実験により評価を行い，それぞれ従来型 HMM と比較して，約 20%と 40%程度の誤り削減率が得られた．順序制約無し AT-HMM において，個々の特徴量の状態共有構造が同一のモデルでは，従来法と比較して認識率が低下しており，順序制約を付けることでより効果的に音声信号がモデル化されたと考えられる．更に AT-HMM のモデル化能力を高めるため，特徴量別音素環境クラスタリング (FW-PEC) の概念を説明し，順序制約付き AT-HMM を生成する特徴量別逐次状態分割法 (FW-SSS) を説明した．切り出し音素認識と孤立単語認識実験により，それぞれ従来法と比較して約 30%と 50%の誤り削減率が得られ，FW-PEC の概念を用いない場合よりも高い誤り削減率が得られた．

今後は，混合分布化した順序制約付き AT-HMM や，不特定話者音声認識，連続音声認識における音声認識性能の評価，更に話者適応法の検討を行う予定である．

参考文献

[1] 松田 繁樹, 中井 満, 下平 博, 嵯峨山 茂樹: “特徴量間で状態遷移が非同期な HMM,” 情報処理学会研究報告, 99-SLP-27-4, 1999-7

[2] S. Sagayama, S. Matsuda, M. Nakai, H. Shimodaira: “Asynchronous Transition HMM for Acoustic Modeling,” Proc. 1999 IEEE Workshop on Speech Recognition and Understanding, to appear in Dec. 1999.

[3] 嵯峨山 茂樹: “音素環境クラスタリングの原理とアルゴリズム,” 信学技報, SP87-86, pp.1-8, 1985.

[4] 鷹見 淳一, 嵯峨山 茂樹: “逐次状態分割法による隠れマルコフ網の自動生成,” 信学論 (D-II), J76-D-II, No.10, pp.1255-1264, 1993-10.

[5] X.D. Huang, K.F Lee, H.W. Hon, M.Y. Hwang: “Improved Acoustic Modeling with the SPHINX Speech Recognition System,” Proc. ICASSP91, pp. 345–348, 1991.

[6] J. Bellegarda, D. Nahamoo: “Tied mixture continuous parameter models for large vocabulary isolated speech recognition,” Proc. ICASSP89, pp.13–16, 1989.

[7] 高橋 敏, 嵯峨山 茂樹: “4 階層共有構造の音響モデルによる音声認識,” 信学論 (D-II), J82-D-II, No.3, pp.315-323, 1999-3.

[8] M. Ostendorf, H. Singer: “HMM topology design using maximum likelihood successive state splitting,” Computer Speech and Language, 11(1), pp.17-41, 1997.