

Buried Markov Model を用いた構音障害者の音声認識の検討*

宮本千琴, 駒井祐人, 滝口哲也, 有木康雄 (神戸大・工), 李義昭 (追手門大)

1 はじめに

近年, 音声認識技術の発展に伴い, 様々な環境下や場面での利用が期待されている. 例えばカーナビゲーションの操作や会議音声の議事録化など様々な分野に応用されている. これらの多くは健常者を対象としており, 文献 [1] [2] では, 構音障害者音声を対象とした特徴量抽出や音響モデル適応, 構築を行っているが, 言語障害者などの障害者を対象としたものは非常に少ない.

言語障害の原因の一つとして, 脳性麻痺が考えられる. 意図的な動作を行う場合や緊張状態にある時に筋肉の制御が難しくなり, 不随意運動を伴う. この運動障害の一つとして, 正しく構音できない場合がある [3]. 本稿では, 脳性麻痺により構音機能に障害を持つ被験者を対象に音声認識実験を行う.

現在の音声認識システムでは, Hidden Markov Model (HMM) が音響モデルとして広く用いられている. HMM は隠れ状態と観測のみからなる単純な構造の確率モデルであり, 各フレームは独立であるという仮定がある. はっきりと発話された音声に対しては高い精度で認識を行うことができるが, 複雑な事象を表現するには適しておらず, 雑音を含む音声や, 連続的に発話された音声を認識する際には, 精度が著しく低下する.

この問題に対し, HMM の独立性の仮定を緩和することでその表現能力を高めたモデルである Buried Markov Model (BMM) [4] が, J. Bilmes によって提案された. HMM の隠れ状態と観測の構造に加え, 各フレームの観測系列間の依存関係をモデルに埋め込んでいるため, より複雑な事象を表現することができる.

本稿では, 発話が不安定な構音障害者の音声認識にこの BMM を適応する. また, BMM の構造を学習する手法として独立性検定を導入する.

2 Buried Markov Model

2.1 概要

Buried Markov Model は HMM の各フレームの観測系列間に依存関係を加えた Fig. 1 のようなグラフィカルモデルによって表現される. BMM の観測系列間の依存関係の構造は, 参照するフレームの隠れ状態に

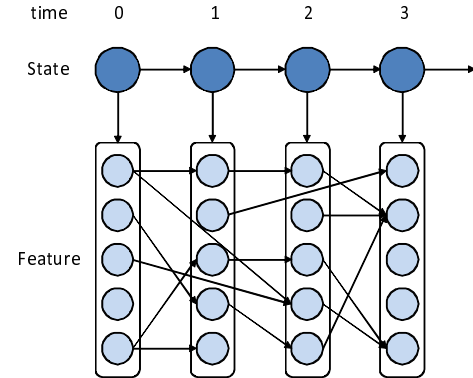


Fig. 1 Buried Markov Model の構造

よって決まるため, 各隠れ状態ごとに学習が行われ, また, 別のフレームからの情報が埋め込まれたモデルを構築することができる.

2.2 構造学習方法

BMM の構造学習には相互情報量を用いた Pairwise アルゴリズムが用いられる. Pairwise アルゴリズムはターゲットノード X に対し, 相互情報量が次の条件を満たすノード Z を親ノード集合 $\mathbf{Z}$ に加える方法である. q はターゲットノードのフレームの状態, $\delta_1, \dots, \delta_3$ はそれぞれの閾値を表す.

$$I(X, Z | q) > \delta_1 \quad (1)$$

$$I(X, Z) < \delta_2 \quad (2)$$

$$I(Z, Z_i) < \delta_3 I(X, Z) \quad Z_i \in \mathbf{Z} \quad (3)$$

式 (1) は相関の高いノードを親とするという条件である. この条件によって, HMM では独立と仮定していた相関の強いノードを特定できる. しかし, 相関の高いノードを親としても, 他の状態においても同様に相関が高ければ, 状態 q を推定するために有益な情報を持ったノードとはいえない. そこで式 (2) の2つの変数間の相互情報量の状態平均が小さいノードを選択する条件を導入し, 候補ノードを状態 q において X との依存性が特異な値をとるノードに限定する. 式 (3) は冗長性の検定を行うための条件である. Z と $\mathbf{Z}$ のノードとの相関が強ければ, $\mathbf{Z}$ のノードから得られる情報と同様の情報しか得られないため, Z を親の候補から除外する.

この手法で用いる相互情報量を計算するために, 初期学習した HMM のパラメータが必要となる. この

*A Study on Dysarthric Speech Recognition using Buried Markov Model, by Chikoto Miyamoto, Yuto Komai, Tetsuya Takiguchi, Yasuo Ariki (Kobe Univ.) and Ichao Li (Otemon Univ.)

方法では HMM のモデリング精度によって得られる依存関係が影響を受ける可能性がある．これを改善するために，HMM の初期学習なしでパラメトリックに BMM の構造を学習するための手法として，独立性検定を用いた構造学習法が提案されている [5]．本稿では，この構造学習法を用いる．詳しくは次章で述べる．

3 独立性検定

3.1 スピアマンの順位相関係数

スピアマンの順位相関係数 (Spearman's rank correlation coefficient) は 2 つの変数の順位の間の相関の強さを表す指標であり，次式で表わされる．

$$\rho = 1 - \frac{6 \sum_{i=1}^n D_i^2}{(n-1)n(n+1)} \quad (4)$$

長さ n の 2 つのデータ系列 X, Y のあるインデックス i におけるデータ X_i, Y_i の順位の差を D_i とする． ρ はデータの順位のピアソンの相関係数を計算したものである．実際に Pairwise アルゴリズムの中で用いる場合， ρ の絶対値が依存性として用いられる．

3.2 ケンドールの順位相関係数

ケンドールの順位相関係数はスピアマンの順位相関係数と同様の 2 つの変数の順位の相関の強さを表す指標であり，次式で表わされる．

$$\tau = \frac{P - Q}{\frac{1}{2}n(n-1)} \quad (5)$$

長さ n の対応したデータ系列 X, Y から 2 つのデータを選ぶ．選んだデータのインデックスを a, b とすると， P は「 $X_a < X_b$ かつ $Y_a < Y_b$ 」または「 $X_a > X_b$ かつ $Y_a > Y_b$ 」の条件に当てはまる組み合わせの数， Q は「 $X_a < X_b$ かつ $Y_a > Y_b$ 」または「 $X_a > X_b$ かつ $Y_a < Y_b$ 」の条件に当てはまる組み合わせの数である．これはすなわち， P が 2 つのデータ間の性の相関性を， Q が負の相関性を表しており， P が大きくなれば正の相関が強くなり τ は 1 に近づき， Q が大きくなれば負の相関が強くなり τ は -1 に近づく． P と Q が同じであれば τ は 0 となりデータ間の相関はないと判断できる．

ケンドールの順位相関係数は，線形・非線形に関わらず相関関係を判別することができるが，複雑な相関関係を判別することができず，単調関数性を判別する指標となる．実際に Pairwise アルゴリズムの中で用いる場合，スピアマンの順位相関係数同様 τ の絶対値が依存性として用いられる．

4 実験

実験用データとして構音障害者 1 名のデータを収録した．発話内容として ATR 音素バランス単語 (216 単語) を使用し，収録は各単語を 5 回連続発声し，その後，各発話を手動で切り出した．収録データのサンプリング周波数は 16 kHz，フレーム窓長は 25 msec，フレーム周期は 10 msec である．このデータを用いて，GMTK[6] によりモデルを作成する．1 回目の発話の認識を行う場合は 2～5 回目の発話を用いて作成し，これを各発話に対して行い，連続音素認識を行う．特徴量には 12 次 MFCC + Δ MFCC の 24 次元を用いる．以上の条件で，構音障害者の音声認識における BMM の有効性を確認する．

5 おわりに

本稿では，発話が不安定な構音障害者の音声認識精度を改善するために，観測ノード間の時間的な依存関係を記述できる BMM を用いた音声認識手法を検討した．

今後は，構音障害者特有の特徴量の検討や，複数話者に対して有効性を確認していく予定である．また，音声特徴だけでなく画像特徴も共に用いた構築アルゴリズムの検討を行う．

参考文献

- [1] 中村 他, “発話障害者音声を対象にした健常者音響モデルの適応と検証,” 音講論 (秋), 109-110, 2005.
- [2] H. Matsumasa *et al.*, “Integration of Metamodel and Acoustic Model for Speech Recognition,” Interspeech2008, 234-2237, 2008.
- [3] S. Terry Canale *et al.*, “キャンベル整形外科手術書 第 4 巻,” エルゼビア・ジャパン, 2004.
- [4] J.A. Bilmes, Buried Markov models: a graphical modeling approach to automatic speech recognition, Computer Speech and Language, Volume 17, Issues 2-3, 213-231, 2003.
- [5] 山本 他, “Buried Markov Model を用いた音声認識モデルの構築法の検討,” 情処研報, 2009-SLP-79, No.21, pp.1-6, 2009.
- [6] J.A. Bilmes *et al.*, The Graphical Models Toolkit: An open source software system for speech and time-series processing, ICASSP, 3916-3919, 2002.