

早稲田大学大学院理工学研究科

# 博士論文概要

## 論文題目

Speech recognition based  
on a Bayesian approach

ベイズ的手法にもとづく  
音声認識

## 申請者

氏名

渡部

晋治

Shinji

Watanabe

専攻・研究指導  
(課程内のみ)

2005年 12月

音声認識技術は音情報をテキスト情報に変換するであり，計算機上での音情報理解の中で最も重要な技術の一つが音声認識技術である．従来の音声認識は研究者間で共通のデータベースを用いることにより限られた環境で性能を競い合うという研究スタイルにより大きく発展を遂げた．しかし，実現されたシステムはモデルパラメータ数百万におよぶ超巨大なものであり，限られた環境に特化されたシステムであるゆえに，頑健性を大きく欠落している．そのため，頑健性の点から音声認識の実用化にまだまだ大きな壁が存在し，いかにして音声認識システムの頑健性を向上させるかは，世界共通の課題となっている．そのような取り組みは音響的及び言語的両視点から研究されており，特に音響的視点においては，モデル化の観点から，従来の隠れマルコフモデル(HMM)にもとづく音響モデルをいかに改良するか，及び，学習理論の観点から，従来の最尤学習にもとづく音響モデル学習をいかに改良するか，という2つの方向性が存在する．本研究は学習理論の観点からベイズ的手法にもとづいた頑健な音声認識実現への取り組みの紹介である．

本論文で扱うベイズ的手法は単に事前確率分布を最尤推定法に取り入れるだけでなく，分布パラメータや隠れ変数といった，モデルのパラメトリック表現において導入された全ての変数を確率変数とみなし，その事後確率分布をベイズの定理から推定して利用する手法であるそのため，従来のパラメータ推定にもとづく最尤推定法とは推定対象をパラメータ値ではなく分布関数とする点が大きく異なる．この事後確率分布推定にもとづいて，ベイズ的手法は，音声認識で広く用いられている最尤学習に比べてより頑健なモデル構築・識別が可能であるといわれている．実際に，ベイズ的手法には大きく分けて3つの利点がある．

- 事前確率分布を介した事前知識の効率的な利用(事前知識の活用)
- モデル構造の多様性を確率変数とみなすことによる，事後確率最大化の意味での与えられた学習データに適したモデル構造の選択(モデル選択)
- モデルパラメータの周辺化による頑健な識別(頑健な識別)

しかし隠れ変数存在下で事後確率分布を正確に推定するためには，モンテ・カルロシミュレーションなどの数値的手法が必要であり，音声認識の音響モデルのような，数百万パラメータが相互に依存し，またHMMや混合ガウス分布モデル(GMM)を通して多数の隠れ変数を内包する複雑なモデルでは莫大な計算量を必要とするため，その実現は大変困難であった．そのため，従来音声認識で実現されてきた事後確率最大化法やベイズ的予測識別法，ベイズ情報量基準法などはいずれも，事後確率分布の推定を行わない近似的実現手法に過ぎず，先に挙げたベイズ的手法の利点を全て内包するものではなかった．

近年，変分ベイズ法にもとづく近似的事後確率分布(VB事後確率分布とよぶ)

推定法が提案され，隠れ変数存在下においても期待値最大化アルゴリズムにより効率よくモデル学習ができるようになった．本研究では，この変分ベイズ法を元に最尤法にもとづく音声認識を発展させ，従来の近似的なベイズ的手法を内包する本格的なベイズ音声認識 VBEC (Variational Bayesian Estimation and Clustering for speech recognition) を構築した．VBEC は大きく分けて 4 つの定式化よりなる．

1. HMM および GMM を用いて表現される音響モデルに対して，出力確率分布とそのモデルパラメータに対する事前確率分布の設定(設定)．
2. 従来の最尤学習法にもとづく Baum-Welch アルゴリズムと同様の VB 版の Baum-Welch アルゴリズムを構築して，モデルパラメータに対する VB 事後確率分布を推定(学習)．
3. 学習データに応じた適切なモデル構造の選択のための VB 評価関数の算出(選択)．
4. VB 事後確率分布と出力確率分布をもとにモデルパラメータに関して周辺化を行い，予測分布が Student の t 分布として解析的にもとまることを示し，その予測分布をもとにした識別(識別)．

これにより，1 から 3 のモデルの設定・学習・選択による音響モデル構築過程及び 4 の識別過程，つまり音声認識の音響モデルに対する全ての過程がベイズ的手法で実現される．従って VBEC は本格的なベイズ音声認識であるといえる．

本研究はこの 4 つの定式化を紹介すると共に，それによって実現されるベイズ法の利点についてそれぞれ実験を用いて検証し，有効性を示す．はじめに，モデルの設定・学習・選択(1 から 3)を用いて，ベイズ的手法を首尾一貫して利用した音響モデル構築を実現する．それにより，少量学習データでの本手法の優位性を示すとともに，音素環境依存の HMM 状態クラスタリング問題及び GMM 混合数決定におけるモデル選択機能効果を示す．次に，上記音響モデル構築過程におけるモデル選択を発展させ，HMM 状態クラスタリング及び GMM 混合数の決定を同時に最適化することにより，音響モデル構造の自動決定を実現する．本手法は，計算機のみによる高性能な音響モデルの自動構築を実現し，さらに従来の人手を用いたモデル構築手法と比較して計算時間を大幅に削減することができる．最後に，モデル識別に注目し，実現される頑健な識別効果を音響モデル適応実験に応用することにより，実用的タスクでの VBEC の有効性を示す．

## 研 究 業 績

種 類 別	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者(申請者含む)
論文 (学術論文)	<b>Shinji Watanabe</b> and Atsushi Nakamura, "Speech recognition based on Student's t-distribution derived from total Bayesian framework," <i>IEICE transaction</i> (掲載決定).
論文 (学術論文)	<b>Shinji Watanabe</b> , Atsushi Sako and Atsushi Nakamura, "Automatic Determination of Acoustic Model Topology using Variational Bayesian Estimation and Clustering for Large Vocabulary Continuous Speech Recognition," <i>IEEE transactions on Speech and Audio Processing</i> (掲載決定).
論文 (英訳論文)	<b>Shinji Watanabe</b> , Yasuhiro Minami, Atsushi Nakamura, and Naonori Ueda, "Selection of Shared-State Hidden Markov Model Structure Using Bayesian Criterion," (English paper of IEICE D-II, vol. J86-D-II, no. 6, pp. 776--786) <i>IEICE Trans. INF. &amp; SYST.</i> , vol.E88-D, no. 1, pp. 1--9, (2005).
論文 (レター論文)	<b>Shinji Watanabe</b> and Atsushi Nakamura, "Acoustic Model Adaptation based on Coarse/Fine Training of Transfer Vectors," <i>情報科学技術レターズ</i> , (2004).
論文 (学術論文)	<b>Shinji Watanabe</b> , Yasuhiro Minami, Atsushi Nakamura, and Naonori Ueda, "Variational Bayesian Estimation and Clustering for Speech Recognition," <i>IEEE transactions on Speech and Audio Processing</i> , vol. 12, pp. 365--381, (2004)
論文 (学術論文)	<b>渡部 晋治</b> , 南泰浩, 中村 篤, 上田 修功, "ベイズ的基準にもとづく状態共有型HMM構造の選択," <i>電子情報通信学会論文誌 D-II</i> , vol. J86-D-II, no. 6, pp. 776--786, (2003). (電子情報通信学会論文賞受賞)
論文 (国際会議)	<b>Shinji Watanabe</b> and Atsushi Nakamura, "Effects of Bayesian predictive classification using variational Bayesian posteriors for sparse training data in speech recognition," <i>Proc. Interspeech '2005 Eurospeech</i> , pp. 1105 1109, (2005)
論文 (国際会議)	<b>Shinji Watanabe</b> and Atsushi Nakamura, "Robustness of acoustic model topology determined by VBEC (Variational Bayesian Estimation and Clustering for speech recognition) for different speech data sets," <i>Proc. Workshop on statistical modeling approach for speech recognition - Beyond HMM</i> , pp. 55--60, (2004)
論文 (国際会議)	<b>Shinji Watanabe</b> and Atsushi Nakamura, "Acoustic model adaptation based on coarse-fine training of transfer vectors and its application to speaker adaptation task," <i>Proc. ICSLP'04</i> , vol. 4, pp. 2933--2936, (2004)
論文 (国際会議)	Parham Zolfaghari, <b>Shinji Watanabe</b> , Atsushi Nakamura and Shigeru Katagiri, "Bayesian Modelling of the Speech Spectrum Using Mixture of Gaussians," <i>Proc. ICASSP'04</i> , vol. 1, pp. 553--556, (2004).
論文 (国際会議)	<b>Shinji Watanabe</b> , Atsushi Sako and Atsushi Nakamura, "Automatic Determination of Acoustic Model Topology using Variational Bayesian Estimation and Clustering," <i>Proc. ICASSP'04</i> , vol. 1, pp. 813--816, (2004)

## 研 究 業 績

種 類 別	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者(申請者含む)
論文 (国際会議)	Parham Zolfaghari, Hiroko Kato, <b>Shinji Watanabe</b> and Shigeru Katagiri, "Speech Spectral Modelling using Mixture of Gaussians," <i>Proc. SWIM</i> , (2004)
論文 (国際会議)	<b>Shinji Watanabe</b> , Yasuhiro Minami, Atsushi Nakamura, and Naonori Ueda, "Bayesian Acoustic Modeling for Spontaneous Speech Recognition," <i>Proc. SSPR'03</i> , pp. 47--50, (2003)
論文 (国際会議)	<b>Shinji Watanabe</b> , Yasuhiro Minami, Atsushi Nakamura, and Naonori Ueda, "Application of Variational Bayesian Estimation and Clustering to Acoustic Model Adaptation," <i>Proc. ICASSP'03</i> , vol. 1, pp. 568--571, (2003)
論文 (国際会議)	<b>Shinji Watanabe</b> , Yasuhiro Minami, Atsushi Nakamura, and Naonori Ueda, "Constructing Shared-State Hidden Markov Models Based on a Bayesian Approach," <i>Proc. ICSLP'02</i> , vol. 4, pp. 2669--2672, (2002)
総説 (招待講演)	<b>渡部 晋治</b> , 南泰浩, 中村 篤, 上田 修功 ``[論文賞受賞記念講演] ベイズ的基準を用いた状態共有型 HMM 構造の選択,' ' 電子情報通信学会技術報告, SP2004-149, pp. 25 30 (2005).
総説 (チュートリアル)	<b>渡部 晋治</b> , ``[チュートリアル講演] ベイズ法を用いた音声認識," 電子情報通信学会技術研究報告, SP2004-74, pp. 13--20, (2004).
総説 (招待講演)	<b>渡部 晋治</b> , ``VBEC:ベイズ的手法にもとづく頑健な音声認識," 音響学会関西支部 第5回若手研究者交流研究発表会, 1-2, 招待講演, (2003)
講演 (学会全国大会)	<b>渡部 晋治</b> , 堀 貴明, ``HMM 状態 - 単語の同時確率を用いた音声言語処理のための複雑度指標," 音響学会講演論文集, 1-5-23, pp. 45-46, (2005.3)
講演 (学会全国大会)	<b>渡部 晋治</b> , 中村 篤, ``移動ベクトルのコース/ファイン学習にもとづく音響モデルの教師付き適応," 音響学会講演論文集, 2-4-11, pp. 107--108, (2004.9)
講演 (学会全国大会)	<b>渡部 晋治</b> , 堀 貴明, Erik McDermott, 南 泰浩, 中村 篤, ``音声認識システム SOLON の日本語話し言葉コーパスにおける評価," 音響学会講演論文集, 2-8-7, pp. 73-74, (2004.3)
講演 (学会全国大会)	<b>渡部 晋治</b> , 佐古 淳, 中村 篤, ``ベイズ的音声認識 VBEC を用いた音響モデル構造の自動決定," 音響学会講演論文集, 1-8-6, pp. 11--12, (2004.3)
講演 (学会全国大会)	P. Zolfaghari, <b>S. Watanabe</b> and S. Katagiri, ``Bayesian Modeling of the Spectrum Using Gaussian Mixtures," 音響学会講演論文集, 2-Q-10, pp. 331--332,
講演 (学会全国大会)	<b>渡部 晋治</b> , 堀 貴明, ``ベイズアプローチによる n-gram 言語モデリング," 音響学会講演論文集, 2-6-10, pp. 79-80, (2003.9).
講演 (学会全国大会)	<b>渡部 晋治</b> , 南 泰浩, 中村 篤, 上田修功, ``変分ベイズ法の音響モデル適応への応用," 音響学会講演論文集, 3-3-12, pp. 127--128, (2003.3).

# 研究業績

種 類 別	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者(申請者含む)
講演 (学会全国大会)	渡部 晋治, 南 泰浩, 中村 篤, 上田修功, ``変分ベイズ法の音声認識への適用," 音響学会講演論文集, 1-9-23, pp. 45--46, (2002.9). (栗屋清学術奨励賞受賞)
講演 (研究会)	堀 貴明, 渡部 晋治, Erik McDermott, 南 泰浩, 中村 篤, ``音声認識システム SOLON の日本語話し言葉コーパスによる評価," 話し言葉の科学と工学ワークショップ講演予稿集, pp.85--92, (2004).
講演 (研究会)	渡部 晋治, 南 泰浩, 中村 篤, 上田修功, ``ベイズ的アプローチに基づく状態共有型 HMM 構造の学習," 電子情報通信学会技術研究報告, SP2002-14, pp. 43--48, (2002)
著書 (国際会議)	Shinji Watanabe, Yasuhiro Minami, Atsushi Nakamura, and Naonori Ueda, "Application of Variational Bayesian Approach to Speech Recognition," <i>NIPS15 MIT Press</i> , (2002).
その他 (学術論文)	Hisakazu Minakata and Shinji Watanabe, "Solar Neutrinos and Leptonic CP Violation," <i>Phys. Lett. B</i> , 468, p. 256, (1999)
その他 (学会全国大会)	渡部 晋治, 中村 博樹, 大場 一郎, ``非断熱的 MSW 機構による太陽ニュートリノフラックスの評価," 日本物理学会 1999 年秋の分科会 (島根大) 1999/09/24a SF-04 (素粒子論)