科学研究費助成事業　　研究成果報告書

平成　２６　年　　６　月　　６　日現在

機関番号：　１３９０３
研究種目：　基盤研究(B)
研究期間：　2009～2013
課題番号：　２１３０００６６

研究課題名（和文）多層モデルの階層間密統合に基づく音声理解フレームワークの研究

研究課題名（英文）Study on spoken language understanding framework integrating knowkedges among multiple layers

研究代表者
李　晃伸（Lee, Akinobu）

名古屋工業大学・工学（系）研究科（研究院）・准教授

研究者番号：８０３３２７６６

交付決定額（研究期間全体）：（直接経費）　　13,500,000円、（間接経費）　　4,050,000円

研究成果の概要（和文）：本研究では，音声認識における信号処理から言語理解までの各層における制約について，低次から高次までの制約を互いに相互作用させる枠組みの研究を行った．階層ごとの統計モデルの研究では，言語・音響・対話の各層における高精度な統計モデルの研究を行い，各層からの制約統合について検討を行った．制約の統合手法の研究では，ベイズリスク最小化探索および対話制御における音声情報の統合等について研究を行った．これらの成果は音声対話システムを構築するための基盤システムとして，オープンソースツールキットMMDAgentおよび音声認識エンジンJuliusの一部として公開されている．

研究成果の概要（英文）：This study focuses on developing a framework that integrates handling of multiple knowledge layer from speech signal processing to spoken language understanding directly into speech recognition process in a statistical mannar. Statistical models at layers of language model, acoustic model and dialogue model are widely investigated. For integration, speech decoding based on Bayes-risk minimization in which all the constraint can be expressed as Bayes risk, and some integration methods that utilizes speech information for dialogue management and turn taking was investigated. Part of the results are publicly available as part of an open-source voice interaction building tool MMDAgent and Julius.

研究分野：　複合領域

科研費の分科・細目：　情報学・知覚情報処理・知能ロボティクス

キーワード：　音声認識　音声言語理解　音声対話　音声信号処理

１．研究開始当初の背景

音声認識・音声言語理解の研究は，大規模な音声データやテキストデータを背景とした統計的アプローチが大いに進展し，実験室環境であれば数十万語の辞書で 95% 以上の認識精度を達成してきた．これは隠れマルコフモデル（Hidden Markov Model; HMM）に基づく音韻モデルや単語 N-gram に基づく言語モデルなどの統計的モデルによる学習手法の進展と，音声やテキストの大量の学習データの整備・収集の飛躍的発達が相互に結びついた結果である．音声認識に直接必要な音韻モデルや言語モデルだけでなく，音声区間検出や話者同定，発話理解や音声翻訳などにも統計に基づく手法が導入・検討されてきた．しかし，いずれも講演音声やニュース等の整った発声を対象としており，人間どうしの日常会話のような話し言葉に対する単語認識精度は依然として 80% にも満たない状況であった．

一般的な音声認識システムは，まず前段の処理として，入力音声に対する雑音抑圧や残響抑圧等の音声信号処理，音声認識に有効な特徴量を算出する特徴量抽出，長時間の入力ストリームから認識対象とすべき区間を同定する音声区間検出 (Voice Activity Detection; VAD) 等の処理を行う．また，認識結果を元に実システムが何らかの処理を行ったり応答を返したりするためには，後段処理として発話理解，談話理解，対話制御，対話管理等を行う必要がある．既存の音声インタフェースや音声対話システムは，各モジュールを段階的に積み上げた構成となっており，これまで個々の階層における技術の精錬・発展が全体としての精度を押し上げてきた．しかし，この垂直な統合の枠組みは問題を各層に分解して解きやすくさせる一方で，日常会話のような話し言葉の，いわゆる ill-formed な音声の場合，前段の誤りが後段の誤りを引き起こし，誤りが増幅されて最終的な認識精度が大幅に劣化してしまう．たとえば，音声区間検出の誤りは，音声が存在しない区間を認識対象としてしまい，認識結果は全て誤りとなってしまう．

われわれ人間は，発話に対して様々な情報から統合的に解釈を行っている．また，例えば「ため息」が音声検出の段階で棄却すべきか，認識対象に組み入れてモデル化すべきか，あるいは言語制約で事後棄却するのか，のように，もともとどの層で扱うべきか曖昧な問題も存在する．このことから，問題を個別の層に分けると同時に，層の間の関係についてモデル化しながら統合されたシステムを構築することが必要である．

２．研究の目的

本研究では，音声認識における信号処理から言語理解までの各層における制約について，低次から高次までの制約を互いに相互作用させ，層ごとの性質を多角的に分析・検討することで，よりよい共通点や相補性を発見して高度につながる枠組みを研究した．研究成果のイメージは，たとえば対話の流れや状況に応じて登場しうる単語を対話管理モジュールが予測しながら，その予測内容から音声認識モジュールの単語の出現確率を変化させるような上位モジュールからのフィードバックを直接認識処理に組み入れるようなことが考えられる．本研究では，そのための統一的な枠組みを実データの検証を通じて検討する．

具体的には，音声認識における各処理層のモデルおよび処理を階層間で密統合する高度な音声理解フレームワークの研究を行う．音声信号処理・音響モデル・発音モデリング・単語モデル・言語モデル・意味解析・対話管理・ユーザモデルといった音声対話のあらゆる層についてその妥当な統計的モデル化と相互の関連性を見い出す．そして総合的に最適化・適応する枠組を提供することで，フレキシビリティを持った音声対話システムの開発基盤を構築・運用する．

３．研究の方法

当該研究目的を達成するための研究計画は，大きく次の３段階に分けられる．

１．階層ごとの統計モデルの研究
２．制約の統合手法の研究
３．基盤システムの構築と応用

まず，第１段階では階層間の密統合に向けて，モジュールごとの統計的モデルについて検討を行う．各層ごとに，高精度な統計モデルに関する研究を行い，各層からの制約統合，あるいは各層へ提供できる制約等について検討を行う．

第２段階では，階層間の制約条件の統合を検討する．各モジュール間で，モデルの条件要素の共通化や依存関係について調べ，統合的なモデルの仕組みを検討し，音声認識において各層の制約を統合する方法を検討する．本研究では各層における制約を直接，音声認識の処理過程に組み入れることで，音声入力に対して全ての層が並列に駆動しながら音高精度かつリアルタイムな音声処理を行う方法を提案する．

第３段階では，階層間の統合手法の枠組みについて検証を行う．各階層ごとのモデルおよび統合手法を実装した音声対話システムの基盤ソフトウェア（ツールキット）を構築し，様々なシステムを試作して検証する．

また，本研究では，研究成果を随時ソフトウェアとして結実させそれを共有化していくことを重要視する．ツールを随時共有化することで，研究の加速，迅速なシステムプロトタイピングが行え，また多様で大規模な実証実験を可能にする．このことは，本研究課題の到達点をより高いレベルに押し上げる

とともに，研究成果を実際に実用化する際の技術上の具体的な道標ともなる．

４．研究成果

前節で述べた研究の各段階ごとについて，研究成果を以下に列挙する．

４．１．階層ごとの統計モデルの研究

研究の方法の第１段階においては各層ごとの統計モデルの研究を行った．得られた成果を以下に列挙する．

（１）言語モデル
国会音声や講演といった自然発話の自動書き起こしの研究において得られた，認識結果の可読性，状況・話題といった知識，あるいは言語・発音モデルの統計的変換手法等の情報を探索過程へ動的に組み込むための制約表現方法について検討した．また言語モデルの話題依存性，話題ごとのコーパス変換やモデル選択，Web 知識を利用した少量テキストからの言語モデルのタスク適応，効率的なワードスポッティング手法についても研究を行った．

（２）音響モデル
多様な話者層や言語に対応する音声認識システムを目指した検討と層間統合について研究を行った．日本語話し言葉コーパス（CSJ)の音響モデルが異なるタスクにどの程度頑健であるか，の調査，多様な話者層や言語に対応する音声認識システムの構築，様々な特徴量やモデルパラメータの動的な制御，話者や環境，コンテキストなどのモデルの制約条件の変化に対応できる音響モデル構造やその計算簡略化，ニューラルネットワークを用いた音声モデルの検討を行った．

（３）対話モデル
統計的ユーザモデル及び統計的対話モデルの研究を行った．ユーザの発話履歴や想定文パターン（文法）とのマッチ度合いを応答候補選択に反映させる手法，ならびに統計的な対話管理手法として近年注目されている POMDP (Partially observable Markov decision process) に基づく音声対話システムの効率性に関する研究を行った．

４．２．制約の統合手法の研究

（１）ベイズリスク最小化探索
ベイズリスク最小化探索による一般的な層間統合の枠組みの提案と検証を行った．音声認識処理において，認識層以外の言語処理・対話・信号処理等の様々な層の制約を，ベイズリスクとして直接組み入れ，その制約を動的に反映した解探索を実現した．研究の初期では，まず音声認識モジュールのみで設定可能なベイズリスクを考え，単語重要度や認識誤りについて有効性のシミュレーションを行った．その後，オープンソースの音声認識エンジン Julius へ手法を組み込み，情報検索システムを構築して検索タスクにおける手法の実験的評価を行った．その後，音声認識モジュールより上位のタスク知識との密統合として，その情報検索システムタスクにおける単語重要度を直接ベイズリスク（誤りリスク）として認識器に組み入れる手法を実装・評価し，有効性を示した．また，誤りリスクの自動決定およびタスク適応についても検証･評価した．
なお，平成 24 年度には，認識エンジン Julius の本体に対してベイズリスク最小化の機能を正式に取り込んで実装し，広く一般公開した．これにより，任意の音声認識システムにおいてベイズリスク最小化探索を実行することのできる基盤を，世の中に広く提供した．

（２）対話制御における音声情報の統合

対話層と他の層との密結合の試行として，発語行為（音声の発話タイミング）を対話制御に直接反映させてモデル化する手法を検討し，評価した．また，対話管理の基礎となるターンテイキングの改善について，ユーザの言い淀みに起因する発話区間の検出誤りから認識誤りや不適切な応答開始を修復する方法を提案した．これは MMDAgent のプラグインとして実装し，デモシステムを公開した．

（３）柔軟な音声認識器のアーキテクチャ

種々の認識システムを柔軟に構築するためのパイプラインを用いたデコーダ実装法について研究を行い，効果的に実装できることを示した．

（４）辞書情報を用いた認識処理の早期確定

言語情報と認識処理の融合の一つとして，辞書情報に基づいて認識結果を早期確定する手法を検証・評価した．また，音声区間検出が音声認識性能に与える影響の調査や，逆に音声認識処理中の情報から音声終了区間を判定（早期確定）する手法の有効性検証など，フロントエンド処理との密統合について研究・検証を進めた．

４．３．基盤システムの構築と応用

（１）基盤ソフトウェア MMDAgent の開発
研究代表者らが構築した音声インタラクションシステム構築ツールキット「MMDAgent」をベースに，実験評価用の音声対話システ

ムを構築した．特に，信号処理から音声認識，応答の表象までを統合した高度なインタラクションを行うためのモーション制御等について考案し実装した．

さらに，汎用対話システム MMDAgent において外部動的情報と連結した動作を記述するための FST 拡張や，様々な情報を扱えるためのシステム改善を行った．

（2）インタフェース

システムの適応の一つとして，性別や年齢層に応じて対話対象のエージェントを切り替えることで対話が活性化することを検証した．また音声入力を用いる UI デザインの指針をユーザビリティ評価により検討した．

（3）データ収集用システムの開発・実験

Android 端末上で動作する音響データ収集プログラム，腕時計型スマートデバイスの検討を行った．また，オンラインで主観評価実験を配布・実行・データ収集するクラウド型音声対話評価実験プラットフォームについて試験実装を行った．

５．主な発表論文等

〔雑誌論文〕（計　７件）

① 南條浩輝，古谷遼，西田昌史，オープンソース音声認識エンジン Julius へのベイズリスク最小化機能の実装と評価，電子情報通信学会論文誌，査読有，Vol.J96-D，No.10，2013，2530--2539

② 古谷遼，七里崇，南條浩輝，音声入力型情報探索におけるベイズリスク最小化音声認識のための単語重要度の自動推定，情報処理学会論文誌，査読有，Vol.54，No.7，2013，1967--1977

③ 秋田祐哉，講演に対する読点の複数アノテーションに基づく自動挿入，情報処理学会論文，査読有，Vol.54，No.2，2013,463--470

④ 鈴木伸尚，西田昌史，山本誠一，文単位で分割されたテキストで学習した言語モデルによる単語信頼度を用いた文境界検出，第 10 回情報科学技術フォーラム(FIT)講演論文集,査読有,第 2 分冊,2011,35--38

⑤ Yuya Akita, Statistical transformation of language and pronunciation models for spontaneous speech recognition, IEEE Trans. Audio, Speech & Language Process.,査読有,18 巻,2010,1539--1549

⑥ Hosan Kamiyama, Takahiro Shinozaki, Koji Iwano and Sadaoki Furui,An Efficient Prosody Application to HMM-based Speech Synthesis, Proc. Asia - Pacific Signal and Information Processing Association (APSIPA)，査読有，1 巻，2010，82-85

〔学会発表〕（計８１件）

① 篠崎隆宏，テーマセッション："「音声認識」は今後こうなる！，SIG-SLP 第 100 回記念シンポジウム（招待講演）2014 年 01 月 31 日～2014 年 02 月 01 日，伊豆長岡温泉 ホテルサンバレー富士見

② 李晃伸，テーマセッション："「音声認識」は今後こうなる！，SIG-SLP 第 100 回記念シンポジウム（招待講演）2014 年 01 月 31 日～2014 年 02 月 01 日，伊豆長岡温泉 ホテルサンバレー富士見

③ Kazunori Komatani, Hierarchical Utterance Understanding for Robust Human-Robot Spoken Dialogues , International Workshop on Spoken Dialogue Systems (IWSDS2014)（招待講演），2014 年 01 月 18 日～2014 年 01 月 20 日，Napa，California，US

④ Kazunori Komatani, Naoki Hotta, Satoshi Sato, Restoring Incorrectly Segmented Keywords and Turn-Taking Caused by Short Pauses, International Workshop on Spoken Dialogue Systems (IWSDS2014)，2014 年 01 月 18 日～2014 年 01 月 20 日，Napa，California，US

⑤ Akinobu Lee, Keiichiro Oura, Keiichi Tokuda , MMDAgent --- A Fully Open-Source Toolkit for Voice Interaction Systems, IEEE ICASSP2013, 2013 年 05 月 26 日～2013 年 5 月 31 日, Vancouver，BC，Canada

⑥ 徳田　恵一，ユーザ参加型双方向音声案内デジタルサイネージシステムの開発・設置・運用事例，日本音響学会研究発表会（招待講演），2013 年 03 月 13 日～2013 年 03 月 15 日，東京工科大学

⑦ 李　晃伸，音声対話システムのさらなる普及には何が必要か，第 95 回音声言語情報処理研究会 SIG-SLP(第 3 回対話システムシンポジウム)パネルディスカッション（招待講演），2013 年 02 月 01 日～2013 年 02 月 02 日，静岡県熱海市

⑧ Takahiro Shinozaki, Pipeline Decomposition of Speech Decoders and Their Implementation Based on Delayed Evaluation, APSIPA Annual Summit and Conference 2012,2012 年 12 月 03 日～2012 年 12 月 06 日，Hollywood，California，US

⑨ Ryuichi Nisimura , Detecting child speaker based on auditory feature vectors for VTL estimation, APSIPA Annual Summit and Conference 2012, 2012 年 12 月 03 日～2012 年 12 月 06 日, 年 12 月 06 日, Hollywood, California, USA

⑩ Yuya Akita, Automatic transcription of lecture speech using language model based on speaking-style transformation of proceeding texts, INTERSPEECH 2012, 2012 年 09 月 09 日～

2012年09月13日, Portland, Oregon, US

⑪ 駒谷和範, 音声対話システム技術の現状と課題, 電気関係学会東海支部連合大会（招待講演）, 2012年09月25日, 静岡大学

⑫ Toshiaki Shimada, Ryuichi Nisimura, Masayasu Tanaka, Hideki Kawahara, Toshio Irino, Developing a method to build Japanese speech recognition system based on 3-gram language model expansion with Google database, ICISS2011 (2011 IEEE International Conference on Intelligent Computing and Integrated Systems), 2011年10月26日, Guilin, China

⑬ Kentaro Suzuta, Ryuichi Nisimura et al., Topic-Dependent Language Modeling for VoiceWeb Systems, WESPAC X 2009, paper-id: 0223, 2009年09月23日, Beijing, China

⑭ Kazunori Komatani, 他4名, Ranking Help Message Candidates Based on Robust Grammar Verification Results and Utterance History in Spoken Dialogue Systems, 10th Annual SIGDIAL Meeting on Discourse and Dialogue, 2009年9月12日, London, UK

⑮ Yuya Akita, Automatic Transcription System for Meetings of the Japanese National Congres, ISCA Interspeech 2009, 2009年9月7日, Brighton Centre, Brighton,UK

⑯ Yuya Akita, Automatic Transcription System for Meetings of the Japanese National Congress, ISCA Interspeech 2009, 2009年9月7日, Brighton Centre, Brighton,UK

⑰ Ryuichi Nisimura , Topic-Dependent Language Modeling for VoiceWeb Systems, HCI International 2009, vol.5611, pp.710-719, 2009年07月22日, San Diego, CA, US

６．研究組織

(1)研究代表者
李　晃伸（LEE, Akinobu）
名古屋工業大学・大学院工学研究科・准教授
研究者番号：８０３３２７６６

(2)研究分担者
駒谷　和範（KOMATANI, Kazunori）
名古屋大学・工学（系）研究科（研究員）・准教授
研究者番号：４０３６２５７９

(3)研究分担者
南條　浩輝（NANJO, Hiroaki）
龍谷大学・理工学部・助教
研究者番号：５０３８８１６２

(4)研究分担者
西村　竜一（NISIMURA, Ryuuichi）
和歌山大学・システム工学部・助教
研究者番号：００３７９６１１

(5)研究分担者
西田　昌史（NISHIDA, Masafumi）
同志社大学・理工学部・准教授
研究者番号：８０３６１４４２

(6)研究分担者
篠崎　隆弘（SHINOZAKI, Takahiro）
東京工業大学・総合理工学研究科（研究院）・准教授
研究者番号：８０４４７９０３

(7)研究分担者
秋田　祐哉（AKITA, Yuya）
京都大学・学内共同利用施設等・助教
研究者番号：９０４０２７４２