

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 5月31日現在

機関番号：34310

研究種目：基盤研究（C）

研究期間：2010～2012

課題番号：22520598

研究課題名（和文） 第二言語話者による英語発話理解技術の研究

 研究課題名（英文） Research on Understanding of English Utterances  
by Second Language Speakers

研究代表者

山本 誠一（YAMAMOTO SEIICHI）

同志社大学・理工学研究科・教授

研究者番号：20374100

研究成果の概要（和文）：本研究ではシステムからの問いかけに対する応答発話を課題文として日本語で表示する translation game 形式により、学習者の応答発話を制約すると共に、学習者コーパスから収集した第二言語話者がおかし易い文法上の誤りを含んだ表現も受理する言語モデルを利用した音声対話型 CALL システムの構成法について検討を行った。具体的な成果としては、①英語母語話者による5段階の主観評価値を付与した学習者コーパスの開発、②それを利用した言語モデルの開発とそれを組み込んだ音声対話型 CALL システムの開発、更に、③課題文の英語表現とは大きく隔たった想定外発話を検出するための手法として、*ngram* と有限オートマトン（FSA）による拘束を実現した言語モデルを用いる音声認識器の音響尤度差、認識結果の編集距離以外に、比較的類似の表現が見受けられる想定外発話をモデル化した想定外発話モデルとの音響尤度差、編集距離を素性として使用する手法の開発である。

研究成果の概要（英文）： In this research we constructed a learner corpus in which each expression was subjectively evaluated by English natives, and we developed a dialogue-based CALL system which showed answers in Japanese to questions from the system to restrict their utterances, using the learner corpus. The CALL system adopts not only commonly used automatic speech recognizers using *ngram* and finite state automata (FSA) designed for accepting utterances in a target domain but also one using a model of out-of-scope utterances. In this research we also propose a novel classification method which classifies the speech recognition result from the recognizer using the FSA into accepted or rejected. The classification method adopts differences among acoustic likelihoods from the multiple recognizers as well as edit distances among the multiple recognizers' outputs as features for the classifier.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010年度	1,400,000	420,000	1,820,000
2011年度	800,000	240,000	1,040,000
2012年度	800,000	240,000	1,040,000
年度			
年度			
総計	3,000,000	900,000	3,900,000

研究分野：人文学

科研費の分科・細目：言語学・外国語教育

キーワード：音声認識、言語処理、CALL、言語モデル

## 1. 研究開始当初の背景

世界のグローバル化に伴い英語の国際標準化は益々進行しており、英語の第二言語話者も種々の場面で状況に適した的確な英語で表現する能力が求められている。音声認識技術を用いた CALL システムは、英語のコミュニケーション能力を効果的に高める手段として期待されており、過去に様々な研究開発が行われてきた。その結果、スクリーン上に提示される予め定められた課題をユーザが発話し、forced-alignment 技術に基づき音響的な特徴や韻律的な特徴を評価するシステム主導型の CALL システムは既に幾つか実用化されている。更に、CMU の FLUENCY のように、学習者の応答発話を数種程度に収めるように対話シナリオを適切に設計し、認識された学習者の応答発話の音響的な特徴や韻律的な特徴を評価するシステムも開発されている。日本人学習者を対象としてこのような複数発話を認識する CALL システムを開発する場合には、文法的な誤りにも対応するように、提示される課題に対応する複数の英語表現に加え、その一部の単語の脱落、置換、挿入により英語母語話者の模範的な表現から逸脱した表現も受理可能とする文法が通常使用される。

今後一層の発展が期待される CALL システムでは、予め設定された課題に対する音響的・韻律的な評価結果のみではなく、ユーザが自発的に発話する表現を認識し、表現上の課題を指摘する機能を備え、外国語学習者にとって理想的な one-to-one interactive language training situations with trained language instructors に類した環境を提供する必要がある。このような機能を備えたシステム・ユーザ協調型の CALL システムでは、想定された表現のみを認識する文法を予め人手で開発して利用する音声認識技術ではなく、多様な表現を認識できる言語モデルを使用する音声認識技術を基本として第二言語話者発話に対する高性能な音声認識を実現することが必要とされる。

第二言語話者の英語は、個々の話者により程度は異なるが、調音は勿論、語彙選択や構文構成においても母語の干渉を広範囲に受ける。このため、確率的な音響モデルや言語モデルを利用する現在の音声認識技術の枠組みの下では、英語母語話者の発話を収集した大規模な音声データやコーパスを用いて学習を行った英語音声認識装置では、第二言語話者による英語への音声認識性能は大きく低下する。この課題を解決するためには、第二言語話者の英語発話および英語表現を収集する必要がある。しかし、音

響モデルに比べて、言語モデルについてはトピック依存性が高いため、トピック毎に言語データを収集してモデルを開発する必要があるが、大規模なコーパスの開発、特に話し言葉のコーパスの開発は英語母語話者のコーパスであっても多大な労力とコストを要する。特に英語を母語としない話者の大規模なコーパス開発は極めて困難であり、第二言語話者の音声認識を困難なものとしていた。

この課題を解決するために、提案者らは、英語母語話者による表現を収集した大規模なコーパスから学習される言語モデル（以下、母語話者モデルと略称）に対し、同一トピックについて収集した日本人の英語表現コーパス（以下、学習者コーパスと略称）を用いて学習される言語モデルの適応化技術の適用を検討する必要がある。

## 2. 研究の目的

本研究では、英語母語話者とは異なる発音の特徴および言語表現としての特徴を有する第二言語話者による自発的な英語発話の高精度な認識を可能とする第二言語話者の英語発話認識技術と、第二言語話者の英語表現をその内容に該当する模範的な母語話者表現への対応付けを行う技術を研究開発すると共に、これらの技術を統合して第二言語話者の自発的な英語発話内容に該当する英語母語話者による模範的な表現を提示する第二言語話者による英語発話理解技術を研究開発し、この技術を基盤としたシステム・ユーザ協調型の CALL システムの実現を目的とする。

## 3. 研究の方法

本研究課題では、①多様な表現の発話を高精度で認識する第二言語話者用言語モデルの構築技術と、②認識された発話をその意図する表現に該当する母語話者表現との対応付けを行う技術の開発を行い、これを用いてユーザが自発的に発話する表現を認識し、表現上の課題を指摘する機能を備えたシステム・ユーザ協調型の CALL システムを開発する。

①については、学習者コーパスでの *uni-gram* 確率、頻度が高く英語母語話者とは異なる特徴を有する冠詞などの特徴的な単位に対する *n-gram* 確率などを拘束条件として言語適応を行う第二言語話者用言語モデルの構築技術について研究を行う。

②については、認識結果を bag-of-words モデルを基盤として Latent Semantic Analysis を用いて次元縮退した空間での距離で、第二言語話者の表現内容を判別する

手法について研究を行う。

#### 4. 研究成果

言語モデルについては、英語母語話者による表現を収集した大規模なコーパス、その一部の表現に対応した学習者コーパスに加えて、英語母語話者による英訳評価が付されている学習者コーパスを開発した。これらのコーパスから統計的な言語モデルと有限オートマトンで表現される言語モデルの開発を行った。認識実験の結果、一定量の表現（本実験では20名の被験者からの発話）を収集した場合、言語的な拘束が強く作用する有限オートマトンによる言語モデル表現の方が高い認識性能を達成することが判明した（他の被験者の自由発話での評価実験で、確率的な言語モデルで74.0%の認識率に対し、有限オートマトンで76.5%の認識率）。すなわち、音声対話型 CALL システム特に translation game 型の CALL システムが対象とするドメインでは、*tri-gram* 程度の拘束条件で言語モデルの適応を実施するより、有限オートマトンのように表現全体に拘束条件が適用される言語モデルの方が現状技術としては適切に機能すると考えられる。有限オートマトンを用いた言語モデルの問題点は拘束条件が確率的な言語モデルより過度に厳しくなる点である。これを解決するために、被験者の発話表現を英語母語話者の主観評価により、誤りは含んでいるが受理可能な表現と、想定外の発話に分類し、受理可能な表現から学習した有限オートマトンの拘束条件を緩和する手法の検討を進め、オープンな発話に対して確率的な言語モデルを上回る認識性能を達成した。

有限オートマトンで認識された認識結果を正しい認識結果かどうかを判別する手法については、有限オートマトンで表現される言語モデルと、対象のドメインを含む大規模なコーパスから学習した統計的な言語モデルとの認識結果の編集距離、音響尤度差に加えて、課題文の英訳文としては不適切であると英語母語話者が判定した表現から学習した言語モデルを構築し、その認識結果との編集距離を使用することにより、認識結果の判別精度を向上させる手法を開発した。

これらの手法を導入した音声対話型 CALL システムの開発を行った。開発したシステムで使用した音声認識エンジンは、通常使用される two-pass 型の認識エンジンであり、実時間認識処理は実現されていない。このため、高速な認識処理を実現し被験者による対話システムとしての効果を検証するために、weighted finite state transducer (WFST) による実現を現在進めている。

#### 5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 5 件）

1. “Gaze and Turn-taking Behavior in Casual Conversational Interactions”, K. Jokinen, H. Furukawa, M. Nishida, S. Yamamoto, *ACM Trans. Interactive Intelligent Systems*, Tiis, Vol. 6 Issue 2, (to publish) (査読有) .
2. “Selection and Rejection of Recognition Results Obtained with Multiple Recognizers of Various Recognition Units Using Support Vector Machine”, N. Suzuki, T. Tanaka, M. Nishida, S. Yamamoto, *Journal of Signal Processing*, Vol. 16, No. 4, pp. 327-330 (2012) (査読有) .
3. “Speaker Clustering in Multi-party Conversation”, M. Nishida, Y. Ishikawa, S. Yamamoto, *Proceedings of the 3rd Nordic Symposium on Multimodal Communication. NEALT Proceedings Series*, Vol. 15, pp.56-61 (2011) (査読有) .
4. “Analysis on the Efficiency of Statistical Measures to Identify Network Structure of Chaos Coupled Systems”, Y. Ueoka, T. Suzuki, and S. Yamamoto, *J. of Modern Physics C*, vol. 21, No. 8, pp. 1065-1079 (2010) (査読有) .
5. “統計的機械翻訳におけるマージン最大化学習” 林克彦、渡辺太郎、塚田元、磯崎英樹、山本誠一、人工知能学会論文誌, Vol. 25, No. 5, pp. 593-601 (2010) (査読有) .

〔学会発表〕（計 22 件）

1. “Differences in Interactional Attitudes in Native and Second Language Conversations: Quantitative Analyses of Multimodal Three-Party Corpus”, S. Yamamoto, K. Taguchi, I. Umata, K. Kabashima, M. Nishida, *Cogsci2013*, Berlin, Germany (to appear).
2. “A Preliminary Investigation on Improving Chinese Pinyin to Character Conversion Using MI Based Automatic Lexical Formation”, J. Zhang , W. Li , X. Wang, M. Nishida, and S. Yamamoto, IEICE Technical Report, SP-2012-98, pp. 1-5, 2013.
3. “Sentence Classification with Grammatical Errors and those Out of Scope of Grammar Assumption for Dialogue-based CALL Systems”, Y. Nagai, T. Senzai, S. Yamamoto, M.

- Nishida, Proc. of *TSD2012*, pp. 616-623, Brno, Czech Republic, Springer.
4. "A Study on Perceptual Training of Mandarin Tone 2 and Tone 3 by Japanese Learners", J. Zhang, Y. Sun, T. Zou, X. Wang, M. Nishida, and S. Yamamoto, IEICE Technical Report, SP-2012-99, pp. 7-12, 2013.
  5. "学習者コーパスを利用した音声対話型 CALL システムにおける想定外発話検出手法", 永井祐, 王晓芸, 西田昌史, 山本誠一, 電子情報通信学会 技報 SP-2012-100, pp. 13-18, 2013.
  6. "An MI Criterion Based Word Segmentation Method and Its Application to Chinese Pinyin-to-word Conversion", J. Zhang, W. Li, X. Wang, M. Nishida, and S. Yamamoto, Proc. 2013 IEICE General Conf., D-14-2, 2013.
  7. "学習者コーパスを利用した音声対話型 CALL システムの開発", 王晓芸, 永井祐, 西田昌史, 山本誠一, 信学会総全大会 D-14-7, 2013.
  8. "Multimodal Corpus of Conversations in Mother Tongue and Second Language by Same Interlocutors", K. Kabashima, M. Nishida, K. Jokinen, S. Yamamoto, *Workshop on Eye Gaze in Intelligent Human Machine Interaction: Eye Gaze and Multimodality*, Santa Monica, Cal., 2012.
  9. "Multimodal Corpus of Multi-party Conversations in Second Language", S. Yamasaki, H. Furukawa, M. Nishida, K. Jokinen, S. Yamamoto, *LREC2012*, pp.416-421, Istanbul, Turkey. 2012.
  10. "Automatic Speech Recognition with Confidence Measures Obtained by Multiple Recognizers of Various Recognition Units", N. Suzuki, T. Tanaka, M. Nishida, and S. Yamamoto, *NCSP2012*, Hawaii, 2012.
  11. "音声対話型 CALL システムでの信頼度評価の検討", 千歳智久, 永井祐, 西田昌史, 山本誠一, 信学会総全大会 2012.
  12. "音声対話型 CALL システムの認識性能評価", 飛田麻夕子, 千歳智久, 永井祐, 西田昌史, 山本誠一, 信学会総全大会 2012.
  13. "Using Features from a Bilingual Alignment Model in Transliteration Mining", T. Fukunishi, A. Finch, S. Yamamoto, E. Sumita, *The 5<sup>th</sup> International Joint Conference on Natural Language Processing*, pp. 49-57, Chiang Mai, Thailand, 2011.
  14. "The Application of Bayesian Alignment Techniques to Transliteration Generation and Mining", A. Finch, K. Yasuda, T. Fukunishi, S. Yamamoto, E. Sumita, *The 5<sup>th</sup> International Universal Communication Symposium, IUCS2011*, Gumi, Korea, 2011.
  15. "A Multimodal Corpus for Modeling Turn Management in Multi-party Conversations", H. Furukawa, M. Nishida, K. Jokinen, S. Yamamoto, *Oriental COCOSA*, Hsinchu, Taiwan, 2011.
  16. "Speaker Clustering Based on Non-negative Matrix Factorization", M. Nishida, S. Yamamoto, *Proceeding of INTERSPEECH2011*, Firenze, 2011.
  17. "Speaker Clustering in Multi-party Conversation and Research on CALL Systems", M. Nishida, S. Yamamoto, *Nordic Symposium on Multi-modal Symposium*, Helsinki, 2011.
  18. "日本語母語話者を対象としたモデル選択と話者適応に基づく英語音声認識", 梶島康輔, 西田昌史, 山本誠一, 信学会総全大会 2011.
  19. "Turn alignment using eye-gaze and speech in spoken interaction", K. Jokinen, K. Harada, M. Nishida, S. Yamamoto, *Proceedings of INTERSPEECH2010*, pp. 2018-2021, Makuhari Messe, 2010.
  20. "Hierarchical Phrase-based Machine Translation with Word-based Reordering Model", K. Hayashi, H. Tsukada, K. Sudo, K. Kevin and S. Yamamoto, *Computational Linguistics*, pp. 439-446, Coling2010, 2010.
  21. "Demonstration of Conversational Eye-gaze Data", K. Jokinen, S. Yamamoto, M. Nishida, *Workshop of Multimodal Corpus of LREC2010*, 2010.
  22. "単語並び換えモデルを考慮した統計的階層句機械翻訳システム", 林克彦, 塚田元, 須藤克仁, Kevin Duh, 山本誠一, 情報処理学会自然言語処理研究会, 2010.
6. 研究組織
- (1) 研究代表者  
山本 誠一 (YAMAMOTO SEIICHI)  
同志社大学・理工学研究科・教授  
研究者番号：20374100
  - (2) 研究分担者

西田 昌史 (NISHIDA MASAFUMI)  
同志社大学・理工学研究科・准教授  
研究者番号：80361442

(3) 連携研究者  
( )

研究者番号：