

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 16 日現在

機関番号：14603

研究種目：若手研究(B)

研究期間：2014～2016

課題番号：26870371

研究課題名(和文) 事象関連脳電位の音声認識フレームワークへの統合のための研究

研究課題名(英文) Integration of Event Related Brain Potentials into Speech Recognition Framework

研究代表者

サクリアニ サクティ (SAKRIANI, SAKTI)

奈良先端科学技術大学院大学・情報科学研究科・助教

研究者番号：30625083

交付決定額(研究期間全体)：(直接経費) 3,000,000円

研究成果の概要(和文)：本研究では、人間の知覚に基づき、脳活動計測の一種である事象関連電位の研究を ASR フレームワークに取り入れ、言語理解における人間の認知プロセスを分析し、ASR フレームワークで統合する可能性を検証しました。ERPによる日本語文での知識と意味のミスマッチ分析について、コミュニケーション不全要因を知覚したときの EEG 信号に差異が生じることが確認された。知識を組み込むために、ASR におけるいくつかのフレームワークを首尾よく調査しました。また、インドネシア、ベトナム、フランスのいくつかの研究機関と共同研究がありました。

研究成果の概要(英文)：Most of ASR systems exist today are still tuned by minimizing word error rate, in which all errors from any words like functional words or fillers are treated in a uniform manner. In fact, the impact of all errors is not the same, but how big and differences are the impact are still unknown. To align with human perception, in this research I propose to utilize the event-related brain potential (ERP) studies into ASR framework, in which the human cognitive process during language comprehension is directly analysis and the possibilities to incorporate within ASR framework are investigated. We were successfully perform the ERP experiments given the ASR results, and we can confirmed that a difference occurs in the EEG/ERP signal when perceiving the communication failure factor. Furthermore, we can learn also how to incorporate different knowledge in ASR framework. We also have opportunities to collaborate with some research institutes from Indonesia, Vietnam, and France.

研究分野：情報学

キーワード：音声情報処理 脳認知科学

1. 研究開始当初の背景

Background

音声認識の研究は過去何十年にもわたり、最初はごく短い音声に反応する単純なシステムから始まり、今では普通の話し言葉でも認識できる機能が携帯電話に内蔵されるほど高機能化している。最新のASRシステムはコーパスベースの統計的手法によるもので、音声信号を統計的アルゴリズムでモデリングしている。その課題は、きれいに発話された読み上げ音声の認識から自然な話し言葉の認識に移行しつつある。したがって、発話内容をそのまま忠実に認識するだけではなく、話者が意図したメッセージを読み取ることが重要である。しかし、現在のASRシステムの多くはまだ単語誤り率 (WER) を下げることに注力している。このWERとは、正しい書き起こしと認識仮説間の最短編集処理(レーベンシュタイン距離)を計算する認識精度の評価測定方法である。このフレームワークでは、キーワードや機能語、あるいは言い淀みやフィラーであってもすべて一様に扱われ、認識間違い(置換、削除、挿入)はすべて同じように「誤り」とみなされる。実際には、こうした誤りの影響がすべて同じであるとはいえないが、その影響の大きさや相違はまだ解析されていない。

ASRシステムの精度測定におけるWERの限界は、これまでに何度か研究対象として取り上げられている。Riccardi et al. 1998の研究では、WERと話し言葉の理解度が比例しないことが証明されている。Wang et al. 2003の研究では、WERと理解度の相違がより顕著に示されている。WERの代わりとなる手法もこれまで多く検証されてきた。Morris et al. 2004の研究では、情報理論を取り入れ、発話内容の自動認識により伝

達もしくは欠落した情報の割合を計算する手法が提案されている。また、Nanjo et al. 2005の研究では、オープンドメインの発話認識向けに新しいASR評価方法およびベイズリスク最小化デコーディングが提案されている。さらに、最新の研究としては、Mishra et al. 2011の提案として、ASR性能の人的主観評価(平均オピニオン評点)により開発された手法がある。しかしながら、これらの研究のいずれも、コミュニケーションの中で人間の脳が言葉をどのように処理し、単語誤りをどのように認識するかを考慮した評価手法には至っていない。

2. 研究の目的

Research purpose

事象関連電位(ERP)の電気生理学的測定は、言語理解の基礎をなす認知プロセスの研究ツールとして大きく期待されている。ERPとは、ある認知事象の直接的結果である脳の電気的な動きを測定したもので、ミリ秒単位の高い時間分解能により脳の活動をオンラインで(刺激を与えた時点のタイミングで)表示することができ、音韻的、意味的、統語的、語用論的情報といった種々な言語情報の瞬時的な認知プロセスにも対応できる。本研究では、人間の知覚に基づき、脳活動計測の一種である事象関連電位

(ERP)の研究をASRフレームワークに取り入れ、言語理解における人間の認知プロセスを分析し、ASRフレームワークで統合する可能性を検証する。

3. 研究の方法

Research method

コミュニケーションにおける単語誤りの影響についてのERP分析のは人間の言語処理は、言語情報の処理と統合を要する極めて

複雑な認知プロセスである。本研究では、
(1) 発話内容：名詞、動詞、形容詞、副詞、
代名詞、およびフィラー、
(2) 誤りの種類：置換、挿入、削除を対象
として意味的および統語的な点での単語誤
りの影響を中心に分析を行う。

研究方法：

- (1) N400やP600成分など、コミュニケーションにおける意味的・統語的単語誤り（名詞、動詞、形容詞、副詞、代名詞の置換、挿入、削除）の影響についてのERP分析。
- (2) ERP結果に基づき、文中の語の重要度係数、意味的係数、統語的係数などASRシステムにとって有用な全知識の抽出・定式化を行うASRフレームワークにおける認知知識の統合を行う。
- (3) ERP結果に基づき、ASRシステムに有用な知識の抽出・定式化を行う。

4. 研究成果

Research achievement

人間の知覚に基づき、脳活動計測の一種である事象関連電位 (ERP) の研究をASR フレームワークに取り入れ、言語理解における人間の認知プロセスを分析し、自動音声認識 (ASR) フレームワークで統合する可能性を検証しました。

- (1) **ERP**による日本語文での知識と意味のミスマッチ分析

特に、本研究では不全要因を

- (a) システム出力が誤りを含んでいた場合
- (b) システム出力は正しいが、コミュニケーションが成立しない場合

この2つの研究をすることができた。

そのうち、前者の例としてASR誤りを、後者の例として未知語の出現を対象とし、それぞれの場面のEEGデータを分析し、コミュニケーション不全要因の検出を試みた。実験の結果、ASR 誤り 時 の ERP (Event-Related brain Potential) 波形から、意味理解の成否に関連してN400 とP600 成分に差異が確認され、構築した識別器では、チャンスレートよりも有意に高い識別精度を示した。また、未知語知覚の脳波実験では、ERP 成分としては未知語知覚時にN400 が、既知語知覚時にはP600 の振幅が有意に増大し、構築した識別器では、チャンスレートよりも有意に高い識別精度を示した。両実験から、コミュニケーション不全要因を知覚したときのEEG 信号に差異が生じることが確認された。

つまり、前者の例としてASR誤りを、後者の例として未知語の出現を対象とし、それぞれの場面のEEGデータを分析し、コミュニケーション不全要因の検出を試みた。

(2) 統計的ASRへの知識の統合

我々は、知識を組み込むために、ASR におけるいくつかのフレームワークを首尾よく調査しました。具体的には、ASR 内のさまざまな深い学習手法を調べました。

我々が研究した技術には：

- ディープニューラルネットワーク (DNN)
- 畳み込みニューラルネットワーク (CNN)
- リカレントニューラルネットワーク (RNN)
- ディープボトルネックニューラルネットワーク (DBN)
- Long Short-Term Memory (LSTM)

我々はまた、これらの技術を組み合わせるための様々なアプローチを検討した。

(3) 共同研究

この研究では、

- インドネシア
- ベトナム、
- フランス

のいくつかの研究機関と協力する機会があります。研究協力だけでなく、共同ワークショップも含まれていました。

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)

Publication (underline your name)

[雑誌論文] (計 件)

Journal

[学会発表] (計 23 件)

Conference

- (1) Andros Tjandra, **Sakriani Sakti**, Satoshi Nakamura, "Compressing Recurrent Neural Network with Tensor Train", Proceedings of The 2016 International Joint Conference on Neural Networks (IJCNN 2017), Alaska (USA), May 2017.
- (2) **Sakriani Sakti**, Seiji Kawanishi, Graham Neubig, Koichiro Yoshino, Satoshi Nakamura, "Deep Bottleneck Features And Sound-Dependent i-Vectors for Simultaneous Recognition of Speech and Environmental Sounds", Proc. IEEE SLT, pp. 35-42, San Diego (USA), Dec 2016.
- (3) Michael Heck, **Sakriani Sakti**, Satoshi Nakamura, "Iterative Training of A DPGMM-HMM Acoustic Unit Recognizer in A Zero Resource Scenario", Proc. IEEE SLT, pp. 57-63, San Diego (USA), Dec 2016.
- (4) Satoshi Tsujioka, **Sakriani Sakti**, Koichiro Yoshino, Graham Neubig, Satoshi Nakamura, "Unsupervised Joint Estimation of Grapheme-to-Phoneme Conversion Systems and Acoustic Model Adaptation for Non-Native Speech Recognition", InterSpeech, pp. 3091 - 3095, San Francisco (USA), Sep. 2016.
- (5) Michael Heck, **Sakriani Sakti**, Satoshi Nakamura, "Supervised Learning of Acoustic Models in a Zero Resource Setting to Improve DPGMM Clustering", InterSpeech, pp. 1310 - 1314, San Francisco (USA), Sep. 2016.
- (6) Quoc Truong Do, **Sakriani Sakti**, Graham Neubig, Satoshi Nakamura, "Transferring Emphasis in Speech Translation Using Hard-Attentional Neural Network Models", InterSpeech, pp. 2533 - 2537, San Francisco (USA), Sep. 2016.
- (7) Quoc Truong Do, Tomoki Toda, Graham Neubig, **Sakriani Sakti**, Satoshi Nakamura, "A Hybrid System for Continuous Word-Level Emphasis Modeling Based on HMM State Clustering and Adaptive Training", InterSpeech, pp. 3196 - 3200, San Francisco (USA), Sep. 2016.
- (8) Hayato Maki, Tomoki Toda, **Sakriani Sakti**, Graham Neubig, and Satoshi Nakamura, "Removing Noise from Event-Related Potentials using a Probabilistic Generative Model with Grouped Covariance Matrices" International Conference of the IEEE Engineering in Medicine and Biology Society, pp.3728-3731, Orlando (USA), Aug. 2016.
- (9) Andros Tjandra, **Sakriani Sakti**, Ruli Manurung, Mirna Adriani and Satoshi Nakamura, "Gated Recurrent Neural Tensor Network", Proceedings of The 2016 International Joint Conference on Neural Networks (IJCNN 2016), Vancouver (Canada), July 2016.
- (10) Michael Heck, **Sakriani Sakti**, Satoshi Nakamura, "Unsupervised Linear Discriminant Analysis for Supporting DPGMM Clustering in the Zero Resource Scenario", International Workshop on Spoken Language Technologies for Under-resourced Languages (SLTU), Procedia Computer Science, Elsevier, Volume 81, 2016, pp73-79, Yogyakarta (Indonesia), May 2016.
- (11) Nurul Lubis, **Sakriani Sakti**, Graham Neubig, Koichiro Yoshino, Tomoki Toda, Satoshi Nakamura, "A Study of Social-Affective Communication: Automatic Prediction of Emotion Triggers and Responses in Television Talk Shows", in Proc. ASRU, Scottsdale-Arizona (USA), December 2015.
- (12) Quoc Truong Do, Michael Heck, **Sakriani Sakti**, Graham Neubig, Tomoki Toda, Satoshi Nakamura, "The NAIST ASR System for the 2015 Multi-genre Broadcast Challenge: On Combination of Deep Learning Systems using a Rank-score Function," in Proc. ASRU, Scottsdale-Arizona (USA). December 2015.
- (13) **Sakriani Sakti**, Faiz Ilham, Graham Neubig, Tomoki Toda, Ayu Purwarianti, Satoshi Nakamura, "Incremental Sentence Compression using LSTM Recurrent Neural Networks", in Proc. ASRU, Scottsdale-Arizona (USA), December 2015.
- (14) **Sakriani Sakti**, Oyunchimeg Shagdar, Fawzi Nashashibi, Satoshi Nakamura, "Context Awareness and Priority Control for ITS based on Automatic Speech Recognition", in Proc. 14th International Conference on ITS Telecommunications (ITST), Copenhagen (Denmark), December 2015.
- (15) Quoc Truong Do, Shinnosuke Takamichi, **Sakriani Sakti**, Graham Neubig, Tomoki Toda, Satoshi Nakamura, "Preserving Word-level Emphasis in Speech-to-speech Translation using Linear Regression HSMMs", in Proc. INTERSPEECH, Dresden (Germany),

- September 2015.
- (16) Hayato Maki, Tomoki Toda, **Sakriani Sakti**, Graham Neubig, Satoshi Nakamura, "Evaluation of EEG Ocular Artifact Removal with a Multi-channel Wiener Filter Based on Probabilistic Generative Model", in Proc. EMBC, Milan (Italy), August 2015
- (17) Hayato Maki, Tomoki Toda, **Sakriani Sakti**, Graham Neubig, Satoshi Nakamura, "EEG Signal Enhancement using Multichannel Wiener Filter with a Spatial Correlation Prior", in Proc. ICASSP, Brisbane (Australia), April 2015.
- (18) Fajri Koto, **Sakriani Sakti**, Graham Neubig, Tomoki Toda, Mirna Adriani, Satoshi Nakamura, "A Study On Natural Expressive Speech: Automatic Memorable Spoken Quote Detection", in Proc. IWSDS, Busan (Korea), January 2015.
- (19) Takafumi Sasakura, **Sakriani Sakti**, Graham Neubig, Tomoki Toda, Satoshi Nakamura, "Unknown Word Detection based on Event-Related Brain Desynchronization Responses", in Proc. IWSDS, Busan (Korea), January 2015.
- (20) Andros Tjandra, **Sakriani Sakti**, Graham Neubig, Tomoki Toda, Mirna Adriani, Satoshi Nakamura, "Combination of Two-dimensional Cochleogram and Spectrogram Features for Deep Learning-based ASR", in Proc ICASSP, Brisbane (Australia), April 2015.
- (21) **Sakriani Sakti**, Yu Odagaki, Takafumi Sasakura, Graham Neubig, Tomoki Toda, Satoshi Nakamura, "An Event-Related Brain Potential Study on the Impact of Speech Recognition Errors", in Proc. APSIPA, Siem Reap (Cambodia), December 2014.
- (22) Fajri Koto, **Sakriani Sakti**, Graham Neubig, Tomoki Toda, M. Adriani, Satoshi Nakamura, "The Use of Semantic and Acoustic Features for Open-Domain TED Talk Summarization", in Proc. APSIPA, Siem Reap (Cambodia), December 2014.
- (23) Fajri Koto, **Sakriani Sakti**, Graham Neubig, Tomoki Toda, Graham Neubig, Satoshi Nakamura, "Memorable Spoken Quote Corpora of TED Public Speaking", in Proc. Oriental COCOSDA, Phuket (Thailand), September 2014.

[図書] (計 件)

Book

[産業財産権]

○出願状況 (計 件)

Patent (in application)

名称 :

発明者 :

権利者 :

種類 :

番号 :

出願年月日 :

国内外の別 :

○取得状況 (計 件)

Patent (obtained)

名称 :

発明者 :

権利者 :

種類 :

番号 :

取得年月日 :

国内外の別 :

[その他]

ホームページ等

Homepage etc.

6. 研究組織

(1) 研究代表者

サクリアニ サクティ (SAKTI, Sakriani)

奈良先端科学技術大学院大学・情報科学

研究科・助教

研究者番号 : 30625083

(2) 研究分担者

()

研究者番号 :

(3) 連携研究者

()

研究者番号 :