

平成30年6月29日現在

機関番号：14603

研究種目：若手研究(B)

研究期間：2015～2017

課題番号：15K16395

研究課題名(和文) 多点表面筋電位信号からの音声認識に基づく発話障害者用意思伝達支援技術の開発

研究課題名(英文) Development of communication aid technology for speech disorder using speech recognition from electromyographic signals derived with multi-electrode array

研究代表者

久保 孝富 (Kubo, Takatomi)

奈良先端科学技術大学院大学・情報科学研究科・特任准教授

研究者番号：20631550

交付決定額(研究期間全体)：(直接経費) 2,500,000円

研究成果の概要(和文)：発話時表面筋電位信号など未知の点が多いデータに深層学習を適用してパターン認識を行うには、ニューラルネットワークの層数・ノード数などを適切に選択する必要がある。平成27年度は入力データに応じてそれらを適切に選択する指標の探索について取り組み、一部を国内外の会議および論文誌にて発表した。平成29年度には時系列データセットに含まれる時間パターンの種類数が未知な場合でも適用可能な分節化手法の開発に取り組んだ。ベータ過程自己回帰隠れマルコフモデルにベイズ階層言語モデルを組み合わせ、その有効性を確認した。本内容についても国際会議・国内会議発表し、論文誌特集号の推薦を受けるなど高い評価を得た。

研究成果の概要(英文)：When deep learning is applied to uncommon data like surface electromyographic signals during a speech, it is necessary to determine parameters, e.g., the number of layers in the network, the number of nodes in each layer, etc. appropriately. In 2015, we developed a method to determine them depending on data, and reported it in a domestic conference, international conference, and a journal. In 2017, we developed a method that can segment time-series data with the unknown number of states by using two Bayesian non-parametric methods, Beta process autoregressive hidden Markov model and Hierarchical Pitman-Yor language model. We reported it in a domestic conference and an international conference. We received a recommendation to a special issue of a journal for our presentation in the domestic conference.

研究分野：医工学

キーワード：時系列分節化 ノンパラメトリックベイズ法 深層学習 表面筋電位信号

1. 研究開始当初の背景

本研究の背景として、発話障害者の意思伝達には非常に多大な困難・苦痛を伴うという問題がある。発話障害者は意識や高次脳機能に必ずしも異常を伴うわけではないことから、発話機能を奪われ他者と意思疎通ができない状況によって強い精神的苦痛が生じてしまう。一方で、意思伝達装置と呼ばれる既存の装置では、(1) 意思伝達の効率が十分改善されていると言えない、(2) 使用方法を学習する負荷が少なからず生じる、といった課題が残されている。これらの課題の克服を目的として、構音障害者の発話動作時に表面筋電位を計測し、得られた電位信号に対して音声認識と同様にパターン認識処理を適用することで発話内容の推定を行う手法が Deng らによって提案された (Deng et al., Proc. Interspeech, 2009)。表面筋電位は筋肉の収縮強度を反映する電気活動の非侵襲的な計測方法である。

発話動作時の表面筋電位信号が、発話した音韻に対して再現性よく同一パターンを示していれば、音声認識と同様に発話内容の推定が可能となる。また、この手法は音声に依存しないことから、顔面・頸部の筋肉の収縮機能が保たれていれば発声量が小さいあるいは発声が全く不可能な場合にも適用可能である。そのため表面筋電位は発話障害者における発話内容推定時の重要な情報源と言える。Deng らの手法は音声認識でも頻用される隠れマルコフモデルに基づいており、その入力として表面筋電位信号を音声信号の代用、あるいは音声信号と併用することで発話障害者の不明瞭な音声でも一定の認識精度が得られた。そして、使用者は意思伝達において従来自然に行う発話動作を行うだけで使用できることから、学習の負荷はほとんど生じない。

上記の様に Deng ら BAE Systems, Massachusetts General Hospital 等からなるグループは、発話障害者を対象に表面筋電位信号を用いた音声認識の有用性を実際に示した。この研究では音素毎のレベルではなく、単語自体をひとまとりの対象として認識している。語彙の向上を図るにはシステムのトレーニングデータとして用いる単語数自体を増やさねばならず、使用者にとって負担となる。他の関連研究として、福田・Bu ら、が、喉頭摘出者 (福田他、電子情報通信学会論文誌, 2005)、および頸髄損傷患者 (Bu, Proc. IEEE Eng Med Biol Soc., 2005) に対して表面筋電位信号から母音認識、あるいはその得られた母音系列からの単語認識を実現した。この手法でも子音に対しての認識は実現されておらず、認識可能な単語数を増やす上では問題となりえる。

発話障害者を対象とした研究は、我々が検索

した範囲で上記のみであるが、健常者を対象とした研究ではカールスルーエ工科大学が他を圧倒している。子音を含めた音素レベルでの認識に基づいた連続音声認識を非発声時においても実現しており (Wand et al., Proc. BIOSIGNALS, 2009; Schultz et al., Speech Comm., 2010)、被験者数についても 70 人以上と大規模なものである。但し、連続音声認識を実現しているとは言え、単語の認識精度は 70%程度に留まっている。近年では、以下で述べる格子状電極を用いての音声認識を行っているが (Wand et al., Proc. BIOSIGNALS, 2013; Wand et al., Proc. IEEE Eng Med Biol Soc., 2013)、この計測手法の導入は我々が先んじて提案したものである。

上記のような先行研究を背景とし、我々は、既存手法で用いられている計測方法に改善の余地があると考えた。まず、電極のサイズが、顔面・頸部などの筋肉のサイズを考慮すると、粗大だと言える。加えて、筋中央付近の神経支配帯と呼ばれる構造を跨ぐような形で既存の皿状電極を配置して表面筋電位を計測した場合、信号の振幅低下が生じてしまうという現象があり、その影響を回避できていない可能性があると考えられた。

我々は表面筋電位からの音声認識に、**図 1**で示したような格子状電極を導入することで、これらの問題の解決を図った (久保他、生体医工学, 2012; Kubo et al., LNCS, 2012)。この電極は、格子点上に多数・高密度の計測点を有しており、上記の 2 点の問題に対して有効な解決策となり得る。実際に我々は、この格子状電極が従来の電極より認識精度向上の上で有効であることを示した。

以上のような内容が本研究の背景として関わっている。



図 1 格子状電極

2. 研究の目的

本研究は発話障害者のための、新たな意思伝達支援技術の開発を目指すものである。我々は音声認識と同様にパターン認識処理を発話中の表面筋電位信号に対して適用することで、そのような支援技術の実現に取り組ん

でいる。我々はこれまでに関連既存研究における計測方法の改善のために多点高密度の計測点を有する格子状電極の導入を提案し、基礎的検証として母音発話課題で従来の電極より高精度の認識が実際に可能となることを示した。本研究では、認識処理部に深層学習およびノンパラメトリックベイズ法と呼ばれる最先端の機械学習手法を導入することによって単語認識を実用レベルの語彙数・認識精度で実現し、発話障害者用意思伝達支援技術の開発に繋がりたいと考えた。

(1) 研究開始当初、優れた性能を示すようになっていた最先端の機械学習手法である深層学習の適用を試みることは関連研究の動向からも重要な試みであると考えられた (Seide et al., Proc. INTERSPEECH2011, 2011; Dahl et al., IEEE Trans. Audio, Speech, Language Process, 2012). (2) 連続音声認識を実現するためには、表面筋電位信号の信号時系列を分節化を行う必要性があった。具体的には、表面筋電位信号を音素単位の区分に分節化可能とするために、ノンパラメトリックベイズ法 (Fox et al., Proc. NIPS, 2008; Mochihashi et al., Proc. ACL-IJCNLP, 2009) の適用を試みようと考えた。

以上の2点に取り組み、発話時の表面筋電位信号のような未知の対象に対しても高精度にパターン認識を行える手法を確立することを本研究の目的とした。

3. 研究の方法

平成 27 年度は、表面筋電位信号に対して深層学習を適用することで発話内容の推定を試みることにした。深層学習は近年の音声認識研究において最も高い認識精度を示しているため、深層学習の導入によって既存手法の推定精度からどの程度の向上が可能であるのか検証を行う。平成 28・29 年度については、単語を発話課題とした実験へと段階を移す計画とした。単語課題時の信号は、前後の音節の影響によって単音節課題の場合よりも非定常な振る舞いを示すと予想された。この前後の音節の影響を回避するためにノンパラメトリックベイズ法を用いて、時系列分節化手法を構築しようと考えた。ノンパラメトリックベイズ法による隠れマルコフモデルの拡張手法は取り得る状態の数を自動的に推定可能な手法であり、前後の音節の影響を克服する上で適しているため、同手法を用いて適切な時系列分節化手法の構築を目指した。

4. 研究成果

本研究では多点表面筋電位信号からの音声認識を実現することで、発話障害者の意思伝達支援に応用可能な技術の確立を目指してきた。この多点表面筋電位信号は格子状の計

測点を有する電極を用いて、そこから表面筋電位を導出することで得られる。この多点表面筋電位信号を発話時に計測し、通常の音声認識と同様にパターン認識処理を施すことで、上記の技術の実現を試みるという方針であった。

平成 27 年度の計画としては、この多点表面筋電位信号の空間パターンに対して、当時様々なパターン認識の応用例で優れた性能を示している深層学習の導入を予定していた。一方で実際に深層学習を適用するに当たり、ニューラルネットワークの層数・層辺りのノード数・適切な事前学習法の選択などについて適切に選択する必要が生じた。平成 27 年度には入力データの空間パターンの複雑さに応じてそれらを適切に選択する指標の探索について取り組み、その一部を国内外の会議にて発表した。

本研究の進捗は初年度時点でやや遅れてしまうようになった。それは、本研究立案・開始後に提案内容と類似した研究内容が発表されるようになり、計画の見直し・再調整などが必要となったためである (Wand et al., Proc. IEEE EMBC, 2014; Diener et al., Proc. IJCNN, 2015)。これらの類似研究に対して優位性・独自性を示せるような計画とするために、初年度の時点で再考の期間を要した。

平成 28 年度は上記の関連内容についてさらに詳細な検討を加えて論文誌への投稿を行い、掲載される運びとなった。また、それと並行して発話時の表面筋電位信号に実際に深層学習の適用を進めたが、上述のように類似研究の報告が増える結果となり、次のステップを優先させることで、他研究と差別化を図れるように計画調整を行った。

最終年度である平成 29 年度には時系列データセットに含まれる時間パターンの種類数が未知な場合でも適用可能な教師なし時系列分節化手法の開発に取り組みだ。具体的には、ベータ過程自己回帰隠れマルコフモデルにベイズ階層言語モデルを組み合わせた手法を提案し、人工データによる検証も踏まえてその有効性を確認した。両手法ともノンパラメトリックベイズ法の一つであり、モデルの複雑さがデータに適応的に調節されるという特徴を有する。提案手法に類似した既存手法として、階層ディリクレ過程隠れマルコフモデルとベイズ階層言語モデルを組み合わせた手法が挙げられるが、提案手法はその手法より、とりわけ時系列データセット内の各時間パターンの出現頻度に不均一性が存在するような場合、優れたパフォーマンスを期待できる。そして、本年度の活動を通じて、時間パターンが複雑なものであっても、それが再現性良く反復して出現していれば、提案手法はその複雑なパターンをシンプルな自

己回帰モデルによって表現される時間パターンを順に組み合わせることで表現した上で、特定できる可能性があることを示した。本内容については国際会議では発表した。また、平成 30 年度に入っても、国内会議での発表を行った。そして、その国内会議では雑誌の特集号への推薦を受けるなど高い評価を得た。また、本内容はその特集号以外の他の論文誌への投稿についても計画している。本研究を適切に用いれば、表面筋電図信号など、潜在するパターン数が未知の信号の時系列データに対しても分節化できる可能性がある。発話時の表面筋電図信号は、単語・音節という階層的な言語情報を反映していると考えられるため、その点でも提案手法のアプローチは有効、適切と考えられる。また、一方で、特段強い制約を仮定している手法ではないので、他分野への応用など汎用性も期待できる。

5. 主な発表論文等

〔雑誌論文〕(計 1 件)

(1) Yasutaka Furusho, Takatomi Kubo, Kazushi Ikeda, Roles of pre-training in deep neural networks from information theoretical perspective, Neurocomputing, 査読あり, 2017, vol. 248, pp. 76-79. DOI: 10.1016/j.neucom.2016.12.083

〔学会発表〕(計 3 件)

(1) Jeric Briones, Takatomi Kubo, Kazushi Ikeda, Double Articulation Approach for Segmentation of Human Interaction using BP-AR-HMM and NPYLM, The Thirty-first Annual Conference on Neural Information Processing Systems (workshop "ADVANCES IN MODELING AND LEARNING INTERACTIONS FROM COMPLEX DATA"), 2017

(2) Yasutaka Furusho, Takatomi Kubo, Kazushi Ikeda, Information Theoretical Analysis of Deep Learning Representations, International Conference on Neural Information Processing, 2015

(3) 古庄泰隆, 久保孝富, 池田和司, ディープニューラルネットワークの入力符号化能力の情報理論的評価, 計測自動制御学会 システム・情報部門 学術講演会, 2015

〔図書〕(計 0 件)

〔産業財産権〕

出願状況 (計 0 件)

取得状況 (計 0 件)

〔その他〕

ホームページ等

<https://sites.google.com/view/milab/home>

6. 研究組織

(1) 研究代表者

久保 孝富 (Kubo, Takatomi)

奈良先端科学技術大学院大学・情報科学研究科・特任准教授

研究者番号: 20631550

(2) 研究分担者

無し

(3) 連携研究者

無し

(4) 研究協力者

無し