前頸部生体信号に基づいた舌運動と黙声単語認識のための基礎研究

—— Recognition of tongue motion and silent speech word based on biological signal from anterior neck ——

2171005

研究代表者　神戸市立工業高等専門学校 電子工学科　准教授　尾　山　匡　浩

［研究の目的］

近年，四肢不随者のためのマシンインタフェースに関する様々な研究が行われてきている。その中で，前頸部から計測した筋活動電位（EMG）を用いて舌動作を推定し，インタフェースとして応用する研究も行われている。舌は残存機能として残っている場合が多く，比較的動きの自由度も高いため，手足に変わるインタフェースとなる可能性があると言える。この舌動作の推定に関しては，顎底部に 20 極程度のアレイ電極を貼り付け，得られた複数の EMG から特徴量を計算し，機械学習手法を用いて 6-10 自由度程度の動作を 90% 以上の精度で推定した報告がされている[1][2]。多極の電極を用いる場合には，装着に時間がかかり，コストも高くなってしまう。関連研究では，舌運動を検出するための電極数や電極配置に関しても検討されているが，更なる検討が必要であるといえる。

一方で，EMG を用いた黙声音声認識に関する研究も行われている[3]。現状，母音の認識精度は高いが，子音では極端に精度が落ちてしまう。また，これらの EMG は口唇周りの表情筋を用いることが一般的で前頸部からの EMG を用いた研究は文献［4］ぐらいである。

そこで，本研究では前頸部から採取された少数の EMG を用い，舌動作の推定と黙声音声認識が可能なシステムの構築を目的とする。

［研究の内容，成果］

（1）舌動作および黙声母音の同時推定

本研究ではまず，舌動作と黙音母音の同時時推定が可能か検討するために図 1 に示すような流れで処理を行った。まず，前頸部から EMG を計測し，標準化を行い，深層学習の 1 つである畳み込みニューラルネットワーク（CNN）を用いて推定を試みた。

（a）EMG の計測と前処理

図 1 に示すように前頸部に 4ch の乾式アクティブ電極を装着し，舌骨上筋群の計測を行った。このとき電極は，喉仏より顎側の前頸部が覆われるように配置した。また 1ch と 3ch に対応する電極ユニットは，幅約 2[mm］の信号線が互いの電極ユニットの間に通るように配置し，2ch と 4ch に対応するものは，左右それぞれの額二腹筋付近に配置した。なお，EMG の計測機器には（有）追坂電子機器の『P-EMG plus』（サンプリング周波数：5000［Hz］）を利用した。

本研究では各被験者において舌の上下・左右・前後の 6 動作を遂行する舌動作タスク時と

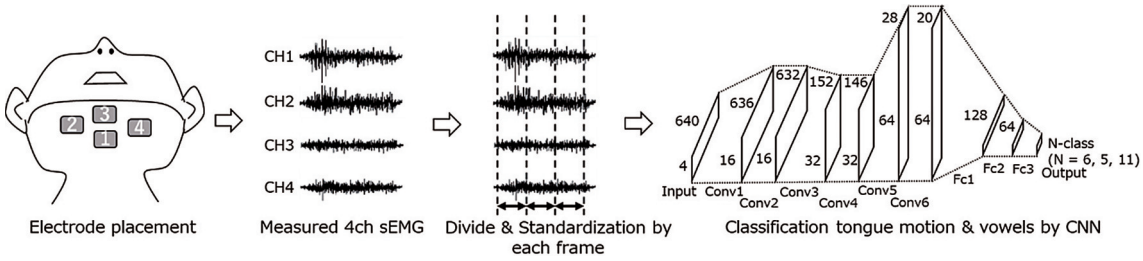

図1　推定手法の流れ

日本語5母音を発声なしに口の形だけ実施する黙音タスク時のEMGを計測した。

計測したEMG信号から，状態の推定を行うため，信号をフレームと呼ばれる一定時間の区間に切り分ける。この時，先行研究を参考にフレーム長を128［ms］（サンプル点数：640点），フレームシフト幅を32［ms］（サンプル点数：160点）として処理する。次に，学習データと評価データそれぞれにおいて，各ch，各サンプル時点で，データ集合の平均と，同様の標準偏差を求め，これを用いて標準化を行った。

(b) CNNを用いた推定

本研究では，近年画像処理分野や音声信号処理分野で目覚ましい成果を上げ，注目されている畳み込みニューラルネットワーク（Convolutional Neural Network：CNN）を推定手法として導入する。CNNは入力信号そのものから学習を行うため，従来のSVM（Support Vector Machine）などのように，前もって特徴量を設計・選択する必要がなく，実験者はCNN構造の設計を行うだけでよい。このため，これまで計測箇所やタスクによって特徴量を決定しなければいけなかったところを，異なる計測箇所や異なるタスクにおいても，全く同じ構造のCNNを用いて統一的に学習と推定を行える可能性がある。そこで本研究では舌骨上筋群のEMGから舌動作と黙声母音の推定を高い認識精度で実現するためにCNNによる手法を導入し，推定を試みる。

CNNは，図1右に示すような構造となっており，全結合層のパラメータ肥大を防ぐため畳み込み層の最終出力サイズが100以下となる範囲で，100［epoch］の学習内で評価データの損失関数の値が充分に収束するように試行錯誤により決定した。

(c) 実験条件

推定手法の有効性を検証するために，実際に計測したEMGを用いて評価実験を実施した。EMGを計測するときには，被験者を椅子に座らせた状態で，タスクについて教示を行い，その後，前頸部の4箇所に電極を装着する。またタスクを実行するタイミングを指示するモニタを用意し，被験者はこのモニタを見てタスクを実行する。各タスクは1動作が約2秒間となるように行ってもらい，それぞれ20動作分計測する。なお被験者は20代男性が4名，30代男性が1名の計5名である。

(d)評価方法

推定は，舌動作タスク（6クラス），黙声タスク（5クラス）に加え，これら2つを合わせたMIXタスク（6+5=11クラス）について，それぞれ行う。また，評価法は（10-Fold Cross Validation：10-FoldCV）を採用し，各クラスの20動作分のうち，18動作を学習データ，2動作を評価データとなるようにデータを分割し，計10組の組み合わせについて推定を行い，その正答率（Accura-cy）を評価値とする。また，CNNによる提案手法の有効性を検証するために従来の特徴量とSVMを用いた推定も試み，それらの比較を行う。特徴量には，EMGを用いた動作推定で広く利用されている

MAV，RMS，WAMP，VAR，ZC，SSC，WL，MMDF，MMNF の 9 種類を用いる。これらの特徴量を 4ch 分算出し，これらを結合することで，9×4＝36 次元の特徴ベクトルを構成し SVM へ入力する。SVM の非線形カーネルには，Gaussian カーネルを用いる。これにより，高次元での線形識別が可能になることが期待される。また，SVM のコストパラメータを C＝10，カーネルパラメータを γ=0.01 として推定を行った。

(e) 推定結果

表 1，表 2 にそれぞれ SVM，CNN を用いた 10-FoldCV における平均の正答率と，その標準偏差を示す。ここで，表中の Accuracy は正答率，SD（Standard Deviation）は標準偏差を表す。

これらの結果をみると，各手法での大まかな正答率の大小関係は変わらず，全体的に CNN が SVM の正答率を上回る結果となった。本研究では舌動作と黙声の同時推定を目指すことから，MIX タスクに注目すると，双方の推定手法において，正答率が最も高い被験者は 1，最も低い者は 3 であった。一方でばらつき（標準偏差）が最も低い被験者は SVM では被験者 1，CNN では被験者 2 であった。

CNN を用いた場合，舌動作は約 87 %，黙声母音は約 84 %，MIX タスクは約 81 % の正答率で推定できている。舌動作の推定に関しては，動作の種類の違いはあるが他の文献では 95 % 以上で推定可能とも報告されている。これらの原因として，EMG の取得 ch 数が少ないことによる情報量の低下，または被験者により同様のタスクを実行する際にも使用する筋やその筋の伸縮度合いが異なることが考えられる。次に，母音の認識に関しても口周りの表情筋を利用した研究では 95 % 以上で推定できており，10 % 以上低い正答率となってしまった。しかしながら，文献［4］の前頸部の EMG を用いた結果と比較すると電極数が少ないにも関わらず，CNN を用いた場合には精度が上回っている。

舌動作に関しては被験者 1 や 4 ではそれぞれ 94 %，98 % と高い正答率の場合もあり，黙声タスクにおいても被験者 1 や 2 が 90 % 以上となっている。このことから，被験者によっては高精度な推定が可能であると言える。

表 1　SVM を用いた推定結果

Accuracy±SD	Tongue motions	Vowels	MIX
Participant 1	84.7±7.1	90.9±7.4	82.8±7.2
Participant 2	78.2±3.6	92.3±7.0	80.8±8.4
Participant 3	69.9±5.7	62.6±8.5	61.0±5.3
Participant 4	96.4±1.6	72.6±4.8	81.4±4.2
Participant 5	65.5±6.3	62.6±8.4	61.5±4.6
Average	78.9±4.9	76.2±7.2	73.5±5.9

表 2　CNN を用いた推定結果

Accuracy±SD	Tongue motions	Vowels	MIX
Participant 1	93.9±5.9	92.1±8.3	91.0±5.7
Participant 2	85.8±4.0	96.5±2.8	89.9±1.9
Participant 3	81.6±5.0	77.5±5.9	67.2±7.1
Participant 4	98.1±0.9	80.1±5.7	86.4±3.2
Participant 5	74.8±5.6	75.8±6.6	71.7±5.6
Average	86.9±4.3	84.4±5.9	81.2±4.7

また，各動作における違いとして，舌動作の「上」と「前」間に相互な誤推定がみられた。また，特に「下」を「後」，「お」を「下」などと，舌の動きが似通っている動作を多く誤推定していることがわかった。これより，舌動作タスク内の「上」，「前」だけではなく，「お」と「下」などの口を閉じて行う舌動作タスクと開いて行う黙声タスクの間においても，舌骨上筋群の筋動作が近いものであれば，誤推定を招く可能性が高いと考えられる。しかしながら，これらの現象がすべての被験者に共通するわけではないこともわかった。また，母音の中では「い」の推定精度が各被験者ともに最も良く，それ以外ではばらつきが見られた。

(2) 黙声単語の推定

次に黙声単語の認識について検証を行った。黙音母音の推定精度では，「い」以外には各被験者間でばらつきが見られたため，今回は先行研究［5］を参考に 4 名の被験者に 20 単語を黙

音発話してもらった際の EMG について計測を行った。なお，1 単語につきおよそ 2 秒間で発話してもらうように被験者に教示し，各単語を 20 回ずつ試行してもらった。

次に，一定の時間幅で信号の切り出しを行う。ここで，フレームの時間幅は動作時間の 2 秒間に加えて，動作前の 0.2 秒，動作後の 0.3 秒を含めた 2.5 秒間とした。また，この切り出したフレームの EMG 信号に対して平均を 0，分散を 1 にする標準化を行った。なお，推定に関しては舌動作および黙音母音の時と同様に CNN を用いて行った。

評価方法については 20 タスク×20 回＝400 のデータ数から 320 の学習データ，80 のテストデータを選び出し 5-fold Cross Validation により 5 つの選び方の組み合わせでの学習・テストをそれぞれ行った。表 3 に推定結果を示す。

この結果より，被験者 3 を除いて 95 % 程度で推定可能であることがわかった。被験者 3 については，舌動作と黙声母音についても良好な結果とは言えないためさらなる検証が必要である。また，先行研究[5]とは利用している筋が異なるが，同程度の推定精度を実現している被験者もいることから前頸部の EMG のみでも黙音単語の識別が可能であると言える。

表 3　CNN を用いた黙声単語推定結果

正答率[%]	被験者 1	被験者 2	被験者 3	被験者 4
Arigatou	100	95	90	100
Gennki	100	95	90	90
Hai	100	95	85	90
Iie	100	100	90	95
Itadakimasu	100	100	90	95
Itamu	100	100	80	80
Kirai	90	100	85	90
Konnbannwa	100	100	60	100
Konnnichiwa	100	100	70	95
Nemutai	100	100	100	100
Nomimono	100	100	80	95
Ohaoyu	100	95	90	95
Okaasann	100	100	75	95
Sayounara	95	100	60	90
Suki	100	100	90	100
Tabemono	100	100	75	95
Toire	100	100	85	100
Tsukareru	90	100	75	90
Uresii	100	100	80	95
AVE.±SD	98.8±3.1	99.0±2.0	82.2±10.7	93.7±5.9

［今後の研究の方向，課題］

本研究では，前頸部に装着した少数の電極から得られる EMG より舌動作と黙声単語の推定が可能か検証を行った。現段階では舌動作，黙音単語ともに先行研究には及ばないが，推定手法の改良や他のセンサと組み合わせることでさらなる向上が期待できる。

また，今後はより多くの被験者から EMG を採取し，リアルタイムで実行可能なシステムを構築することで，被験者がシステムに十分に慣れた場合の検証も行いたいと考えている。加えて現在，コンパクトでより安価な EMG のアクティブ電極回路と治具を試作しており，それらを用いた検証実験も実施していきたい。最終的には舌動作による方向指示と黙声単語による命令指示を組み合わせたヒューマンインタフェースを実現したい。

［成果の発表，論文等］

① 渡邉　大生，尾山　匡浩，福見　　稔，“舌骨上筋群の表面筋電位に基づく CNN を用いた舌動作と黙声の推定”，平成 29 年電気学会電子・情報・システム部門大会，pp. 1553-1554, 2017

② 尾山　匡浩，渡邊　大生，“表面筋電位を用いた舌動作および黙声認識”，第 18 回計測自動制御学会システムインテグレーション部門講演会, pp. 2650-2651, 2017

③ 尾山　匡浩，渡邊　大生，“EMG に基づく舌動作と黙声母音の識別に関する検討”，信学会福祉情報工学研究会（WIT2018-4), pp. 17-20, 2018

④ 渡邉　大生，尾山　匡浩，福見　　稔，“舌骨上筋群の筋電信号に基づく CNN を用いた舌動作と黙声母音の推定”，電気学会論文誌 C, Vol. 138, No. 7, pp. 828-837

［参考文献］

［1］長谷川ら “SVM を用いた前頸部生体電位信号に基づく舌運動の検出”，日本機械学会論文集, Vol. 78, No. 796, pp. 3970-3978,（2012）

［2］ M. Ssaki et. al., “Tongue interface based on surface EMG signals of suprahyoid mus-cles”, ROBOMECH Journal, 3 (1), 9, (2016)
［3］真鍋ら“無発声音声認識：筋電信号を用いた声を伴わない日本語 5 母音の認識”, 信学論 D-II, Vol. J88-D-II, No.9, pp. 1909-1917, (2005)
［4］T. Kubo et. al., “Towards excluding red-undancy in electrode grid for automatic spe-ech recognition based on surface EMG”, Neurocomputing, Vol. 134, 15-19, (2014)
［5］福田ら, “EMG 信号を利用した代用発声システム”, 信学論 D-II, Vol. J88-D-II, No. 1, pp. 105-112, (2005).