

舌骨上筋群の EMG を用いた無発声単語推定 Word Estimation of Silent Speech Using EMG from Suprahyoid

尾山 匡浩[†] 渡邊 大生[‡]
Tadahiro Oyama Taisei Watanabe

1. はじめに

筋萎縮性側索硬化症(ALS)や四肢不随の人々は、他人や機械に意思を伝達することが容易ではない。そのため、そのような方々のための human machine interface(HMI)の研究開発が行われてきている。代表的な例として、脳波を用いた brain computer interface(BCI)や視線によるものがあげられる。その中の 1 つとして、筋活動電位(EMG)を用いた黙音認識に関する研究も行われている[1-4]。

例えば、[2]や[3]では口唇周辺の EMG を利用して日本語母音の推定を試みている。また、福田らは日本語 20 単語に対して認識実験を行っており、95%程度の精度で推定可能と報告されている[4]。このように黙音を EMG から認識する場合には、顔の口唇周辺に配置した電極から得られる EMG を利用することが多い。

一方で、Kubo らは口唇周辺ではなく前頸部から計測された EMG のみを用いた黙音推定を試みている。この研究では、日本語 5 母音を 80%程度の精度で推定が可能であると報告されている[5]。また、我々も同様に前頸部に少数の電極を配置し、そこから得られる EMG を利用して舌動作と黙音母音の認識手法について検討してきた[6]。しかし、実際に利用する場面を想定した場合には、母音ではなく、簡単な単語(例えば、ご飯、トイレなど)を認識可能な命令コマンドとしてシステムに登録しておく方が、より使いやすい HMI になると考えられる。

そこで本研究では、前頸部から計測した EMG から複数の黙音単語の推定が可能か検討する。本稿では、推定手法の提案と実際に計測した EMG に対して行った推定実験結果について報告する。

2. 提案手法

2.1 提案手法の流れ

本稿で提案する手法を図 1 に示す。まず、前頸部から得られた EMG 信号をフレームとして切り出し、各チャンネル・各時点のデータ集合全体で標準偏差が 1 になるよう揃える標準化を行う。その後、特徴抽出を行い、support vector machine(SVM)により単語の推定を行う。次節以降に詳細な説明を示す。

2.2 データ計測・前処理

本研究では、EMG の計測に(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、ハードウェアにて 1000 倍の増幅と帯域幅 1.64[Hz]~400[Hz]の BPF をかけるよう設定した。電極は付属の乾式のアクティブ電極ユニットを 4 個、湿式電極(リファレンス電極)を 1 個用いた。乾式電極ユニット(縦約 2.5[cm]、横約 1.8[cm])は図 1 に示すように前頸部に配置し、湿式電極は

その大きさ(直径約 3.0[cm])から今回は右手甲に装着する。ここで図中の数字は各電極ユニットに対応する ch 番号を示している。ここで、すべての電極ユニットにより、喉仏より顎側の前頸部が覆われるように配置する。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。

このようにして EMG を計測した後、一定の時間幅で信号の切り出しを行う。本稿の実験では、後述するように 1 単語を 2 秒間で発話してもらうように被験者に教示したため、2.5 秒間の EMG 信号を 1 単語分のフレームとして切り出しを行う。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 秒を含めているため 2.5 秒間となっている。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 にする標準化を行う。

2.3 特徴抽出・識別

EMG での特徴量には、EMG を用いた動作推定で広く利用されている mean absolute value (MAV), root mean square (RMS), willison amplitude (WAMP), variance of the EMG (VAR), zero crossing (ZC), slope sign changes (SSC), waveform length (WL), modified median frequency (MMDF), modified mean frequency (MMNF)の 9 種類[7]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、 $9 \times 4 = 36$ 次元の特徴ベクトルを構成する。

特徴量からの単語推定には SVM を利用する。本稿では非線形カーネル関数としてガウシアンカーネルを用い、またハイパーパラメータ(コストパラメータとカーネルパラメータ)は Grid-Search により決定する。

3. 検証実験

3.1 計測条件

計測時には、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う。その後、前頸部の 4 箇所(2 対)の電極を、右手背にリファレンス電極を装着する。また、EMG 波形を確認するモニタと、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、被験者はこれを見てタスクを実行する。表 1 に本研究にて用いる 20 個のタスクの詳細を示す。このタスクは、文献[4]と同様であり、被験者はこの 20 単語を 1 単語につき 20 回ずつ黙音で発話(声を出さずに口の形のみ変形)する。具体的なタイムチャートとしては、まず 'Arigatou' において、休み 2 秒・実行 2 秒(この 2 秒の間に 1 単語の黙音発話を終える)の 4 秒を 20 回繰り返し行う。次に実行する黙音単語を 'Gennki' に切り替え同様に 20 回繰り返し、これを最後の 'Ureshii' まで続ける。なお、被験者は 20 代男性が 3 名、30 代男性が 1 名の計 4 名である。

[†] 神戸市立工業高等専門学校, KCCT

[‡] 奈良先端科学技術大学院大学, NAIST

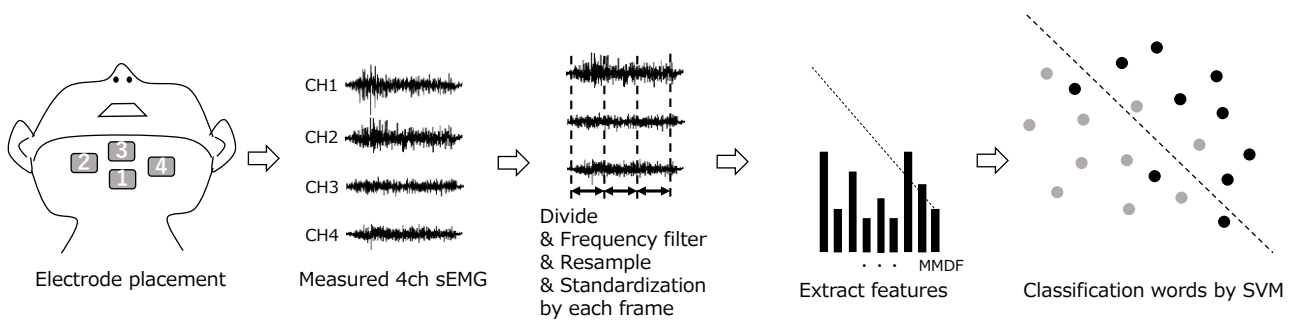


図 1 提案手法の流れ

表 1 タスク(黙声単語)表

Task	タスク	Task	タスク
Arigatou	ありがとう	Nomimono	のみもの
Gennki	げんき	Ohaoyu	おはよう
Hai	はい	Okaasann	おかあさん
Iie	いいえ	Onegai	おねがい
Itadakimasu	いただきます	Sayounara	さようなら
Itamu	いたむ	Suki	すき
Kirai	きらい	Tabemono	たべもの
Konnbanwa	こんばんは	Toire	といれ
Konnichiwa	こんにちは	Tsukareru	つかれる
Nemutai	ねむい	Ureshii	うれしい

表 2 各被験者の黙声単語推定結果[%]

Task	被験者 1	被験者 2	被験者 3	被験者 4
Arigatou	95	80	75	95
Gennki	90	65	65	65
Hai	75	75	55	40
Iie	75	45	40	40
Itadakimasu	80	40	40	60
Itamu	70	100	5	45
Kirai	40	80	50	65
Konnbanwa	80	60	15	85
Konnichiwa	45	55	20	40
Nemutai	50	80	50	80
Nomimono	65	80	10	60
Ohaoyu	85	70	30	65
Okaasann	55	70	10	60
Onegai	85	85	25	70
Sayounara	70	30	0	60
Suki	85	55	25	75
Tabemono	40	70	50	70
Toire	75	50	60	60
Tsukareru	65	45	20	55
Ureshii	85	85	40	95
AVE.±SD	70.5±16.7	66.2±17.5	34.2±21.4	64.2±16.3

3.2 評価条件

推定は、単語タスク (20 クラス) について、それぞれ行う。20 タスク×20 回の合計 400 個のデータから 320 個の学習データ、80 個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、学習・評価をそれぞれ行う。評価値の指標として、本稿では正答率 (Accuracy) を利用する。

3.3 実験結果

表 2 に提案手法を用いて推定を行った際の各被験者の正答率を示す。ここで、表中の最下段は平均正答率と標準偏差を表している。被験者 3 を除いて、65% 程度で推定できているが、文献[4]と比べると非常に低い結果となっている。この原因として、EMG の計測位置の違いや SVM のパラメータの探索が十分ではない可能性が考えられる。

4. おわりに

本稿では、前頸部から計測された EMG を用いて黙声単語の推定を試みた。今後は、深層学習の導入など推定手法の改良や推定結果の詳細な検証が必要であるといえる。

謝辞

本研究は、公益財団法人立石科学技術振興財団の助成を受けたものであり、ここに謝意を表す。

参考文献

- [1] B.Denby, T.Schultz, K.Hond, T.Hueber, J.M.Gilbert, J.S.Brumberg, "Silent speech interfaces", Speech Communication, vol. 52, Issue 4, pp.270-287, (2010).
- [2] 真鍋 宏幸, 平岩 明, 杉村利明, "無発声音認識: 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).
- [3] 村上 弘樹, 村治 雅文, 白藤 立, "筋電信号によるメンバシップ関数を用いた日本語単音の認識", 電気学会論文誌 C, Vol.136 No.12 pp.1821-1826, (2016).
- [4] 福田 修, 藤田 真治, 辻 敏夫, "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).
- [5] Takatomi Kubo, Masaki Yoshida, Takumu Hattori, Kazushi Ikeda, "Towards excluding redundancy in electrode grid for automatic speech recognition based on surface EMG", Neurocomputing, Vol.134, 15-19, (2014).
- [6] 渡邊 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, (to appear)
- [7] Angkoon Phinyomark, Chusak Limsakul, Pornchai Phukpattaranont, "A Novel Feature Extraction for Robust EMG Pattern Recognition", JOURNAL OF COMPUTING, Vol.1, No.1, pp, 71-80, (2009).