

深層学習を用いる口唇 EMG によるサイレント音声認識と個人認証

代表研究者 福見 稔 徳島大学大学院社会産業理工学研究部 教授

1 はじめに

本研究では、深層学習ニューラルネットワークと統計学習法を用いて、口唇周囲で計測した EMG（筋肉電位）を用いるサイレント音声認識（無発声音声認識）および個人認証を高精度に実現することを目指した。

音声は人間にとって重要なコミュニケーション手段のひとつである。しかし、喉頭癌などにより喉頭全摘出やストレス性障害が原因で後天的に発声機能障害を発症する患者が存在する。発声機能障害を発症した場合、自力での発話は困難となり日常生活において大きな制限を負う。そのため、発声機能障害の患者は代替手法を使用することでコミュニケーションを行うしかなく不便な生活となっている。主要な代替手法としては笛式人工喉頭をはじめとする外科的処置を必要とする方法[1]、食道発声法をはじめとする人間が持つ本来発声に使用しない部分を使用する方法[1]、電気式人工喉頭などの機器を使用する方法[1]が存在する。しかし、外科的処置を必要とする方法は文字通り外科的な処置が必要であり、本来発声に使用しない部分を使用する方法は長い訓練期間を必要とし、機器を使用する方法は一般に機器が高価であり、使用の際に片手を使うといった身体的な制限も必要となるなどの問題点がある。

最近、明瞭な音声を得ることができない状況における音声認識や発声補助といった研究[2][3]が盛んに行われるようになった。これらの研究は主に携帯機器や PC などのキーボード操作、フリック操作が困難な状況での利用が想定されており、口唇周辺の画像による視覚認識だけで発話を認識する試み[2]や表面筋電信号を利用した手法などが検討されている[3]。

本研究では、口唇周囲の EMG 信号（表面筋電信号：Electromyogram）に着目してサイレント音声認識と個人認証を行うことが目的である。通常のサイレント音声認識では母音認識は比較的に高精度に認識可能であるが子音認識は精度が低い状態が続いている。また、従来の研究で使用する EMG 計測センサは湿式が多く、乾式センサの例では認識精度が低く実用性が低い。本研究では特に乾式センサを使用して、まず母音認識での高精度化と個人認証を目指す。母音認識と個人認証で高精度化を実現できれば、さらに子音認識での高精度化を目指したい。

本研究の成果により、騒音下や静寂さが求められる場所での無発声会話の実現、また喉頭切除や発声機能障害により発話が困難な場合においては、新しい入力インタフェースや代用発声システムとして有効活用可能となる。さらに、音声を使用せずに口唇 EMG による個人認証（偽造は困難）という新しい技術分野を開拓でき、サイレント音声認識の安全性を高めることが可能となる。

これらにより、従来、遠隔医療や携帯機器の利用などを阻害していた要因を削減し、電気通信技術・分野の利用拡大に繋がり、電気通信技術による障がい者福祉にも寄与できる。

本研究では、その目的のために機械学習手法である深層学習とサポートベクタマシン(SVM)を用いている。

2 EMG 計測センサと機械学習法

以下では、EMG 信号の概略とその計測センサについて説明し、さらに本研究で使用する機械学習法について簡単に説明する。

2-1 EMG 信号[4]

生物が運動を行っているとき、その基となる駆動力は骨格筋の収縮によって生み出される。この筋活動は脊椎の中にある α 運動ニューロンの興奮から始まると考えられている。 α 運動ニューロンが脳からの指令や脊椎を通る種田根の反射によって興奮すると、そのインパルスが神経軸索を經由して目指す筋に伝えられる、ひとつの運動ニューロンから生じる神経軸索は筋の中で枝分かれすることで多数の筋繊維に神経筋接合部を作る。運動神経の興奮は神経インパルス列として神経筋接合部に到達する。神経筋接合部では神経終末から化学伝達物質であるアセチルコリンが放出され筋繊維側の電気的な興奮、筋繊維膜上の脱分極を生じさせる。筋繊維膜上の興奮は筋小胞体からのカルシウム放出を引き起こし筋繊維内に存在するミオシン分子とアクチン分子の相互作用が可能となり、ミオシン繊維とアクチン繊維の滑り運動が生じることで筋張力が発

生する[4]。筋電図には筋収縮のレベルに応じて発生する動作や力の入れ具合の情報が含まれている。また、筋繊維の電気的興奮は筋繊維の両端へ向かって3~6m/secの速度で伝播するが電気的興奮と共に筋収縮は起こらず30~100ms程度の遅延が発生する。このことから筋電信号は運動動作よりも先行して発生しているため筋電信号を計測することで動作の推定が可能であると考えられる。

2-2 EMG 計測センサ

表面筋電を計測するための電極は大きく2種類に分けられる。湿式電極と乾式電極である。湿式電極(図1)は心電図計測などに利用されるものと同様の医療計測用である。電極と皮膚表面間の接触インピーダンスを下げるために、使用する際には貼り付け部位の準備や電極ペーストを必要とする。これは高い接触インピーダンスのままでは湿度や温度、動作の影響を受けることで分極電圧が変化し、動作の激しい運動時に発生するアーチファクトを回避するためである。しかし、貼り付けや取扱いに手間がかかり、電極ペーストの乾燥などの劣化により長期間の使用が難しいため高コストとなっている。加えて、被験者によってはアレルギー問題を引き起こす可能性がある(一定割合の被験者が該当する)。しかし電極インピーダンスが低いために安定した表面筋電位を測定することができる点が最大のメリットである。一方、乾式電極(図2)は貼り付けや扱いは容易である反面、湿式電極と比較して密着性が低いことからアーチファクトが発生しやすく、EMG信号の検出性が低下する問題がある。本研究では将来的な実用性を考慮し、あえて乾式電極を用いて研究を進める。図2では、3極の電極が示されている。3極の内、真ん中の電極がレファレンス(アース)電極である。



図1 湿式用ディスプレイ電極。



図2 乾式電極(3極)。

2-3 機械学習法

本研究では、口唇周囲のEMG信号の認識に機械学習法である深層学習ニューラルネットワーク[5]とサポートベクタマシンを用いる。これらについて簡単に説明する。

ニューラルネットワーク(Neural Network, NN)は生物の脳神経系を模した情報処理システムであるニューロンモデルを多層構造にしたものである。ニューロンモデル一つは比較的単純な非線形素子であり、そのシナプス結合部分を重みとして学習可能となっている。したがって、3層以上の多層ニューラルネットワークは非線形な関数を近似可能であり、学習により万能性を発揮する。次に、深層学習とは従来の3層程度であったニューラルネットワークを4層以上に拡張した深い層を持つニューラルネットワークである。2006年以降に北米を中心として深層構造のニューラルネットワークの学習法が次々と開発され、第三次人工知能ブーム(同時に、ニューラルネットワークブーム)となっている。現在、世界中で人工知能関係のブレイクスルーが起こり、画像認識、音声認識などで従来のアプローチを大きく凌駕する性能を発揮している[5]。

深層学習の中でも画像認識等でブレイクスルーを発揮しているのが、たたみ込み層を有する構造のニューラルネットワーク(CNN)である。たたみ込み層は、日本で開発された“ネオコグニトロン”[6]の学習を勾配法に置き換えたものとなっている。このネオコグニトロンを参考にして、勾配法に基づくたたみ込み演算とプリーング[7]が使用されるようになった。ただし、詳細な説明は紙数の都合で割愛する。

本研究では、口唇周囲のEMG信号の認識に対して、深層学習ニューラルネットワークを適用する。特に本研究の実施期間内では、日本語5母音の識別に深層学習を用いている。なお、研究期間の制限もあり、深層学習を個人認証に用いる方法は予備的な適用実験に止まっている。

次にサポートベクタマシン(SVM: Support vector machine)について簡単に説明する。SVMはクラス分類、回帰、新規性検出などの分野で近年よく使われるようになった手法である。SVMの特徴はモデルパラメータがある凸最適化問題の解として求まるため、局所解があればそれが大域解にもなる点である。SVMでの識別対象は通常2クラス分類問題である。この2クラスの間にはマージン最大化を考慮して、分類に最適なクラス境界を決めることが目的となる。一般的に、SVMの最大の特徴はマージンの最大化にあると言われているが詳細は割

愛する。

3 EMG 信号による 5 母音識別と個人認証

EMG 信号の識別のために、本研究での処理の流れは3つに分類される。最初はデータ計測による入力部、次にデータ変換部、さらに学習識別部である。深層学習の場合、学習部分に特徴抽出を含む仕組みとなっている。従来の方法（例えば SVM）では、特徴抽出部と学習識別部は別々になっていた。そのため、SVM の学習を行う際は、FFT スペクトルに対して判別分析（LDA : Linear Discriminant Analysis）を適用し、次元圧縮を行った。以下では、これらについて簡単に説明していく。

3-1 入力部

入力部では 2 極の小型乾式電極を 3 個用いて EMG 信号を計測する。電極は口輪筋、大頬骨筋、口角下制筋の信号が計測できる場所に装着し、サンプリングレート 1,000 [Hz] で計測を行った。計測機器は追坂電子機器製 Personal-EMG plus [8] (図 3) を用いた。計測機器本体はすでに研究室に備わっていたが、小型の 2 極乾式電極 (図 4) は、本研究助成により 3 個購入した。また、計測機器側でハムフィルタ、および 5 [Hz] ~ 500 [Hz] のバンドパスフィルタを適用した。小型の電極であるため皮膚との接触面積が小さく、構造上、ノイズが重畳しやすくなっている。また、小型乾式センサの場合には、レファレンス電極を別途使用する必要がある。

本研究においては 3 つの二極式乾式電極を使用して表面筋電を測定する。日本語 5 母音は発声時の口唇形状において顕著な特徴が現れる。表 1 に各母音発声時の下頬骨、口唇、口角の状態を示す。これより、母音を測定するためにはこの 3 部位の情報を得ることができる部位に電極を設置するのが適当であると推測できる。よって本研究では 3 つの電極を口輪筋 (1ch)、大頬骨筋 (2ch)、口角下制筋 (3ch) の情報が測定可能な場所に設置した (図 5)。口輪筋の特徴は口唇の縮小と収縮であるために全母音において特徴的な筋電位が、大頬骨筋には口角を外上方に引くため、母音「う」及び「お」に対して有効な筋電位が、そして口角下制筋には口角を下方に引く作用があるため、母音「い」及び「お」において有効な筋電位が取得可能であると考えられる。

3-2 入力部

データ変換部では取得したデータに FFT (高速フーリエ変換) を適用しパワースペクトルに変換する。その後、ノイズ除去などの処理を行い、識別に使用するためのデータセットの作成を行う。

本研究においてサンプリングレートは 1,000 [Hz] である。また、先行研究において、開口部に多数のノイズ成分の混入が確認できた [9]。これより、本研究においては長音部のみを使用し、データセットを作成した。取得したデータの中心を基準として、前後 2,048 点、合計 4,096 点を抽出する (図 6)。図 6 で横軸が計測時間、縦軸は信号のパワーを示している。



図 3 Personal EMG plus (追坂電子器機製)。



図4 2極式小型乾式センサ（迫坂電子器機製）.

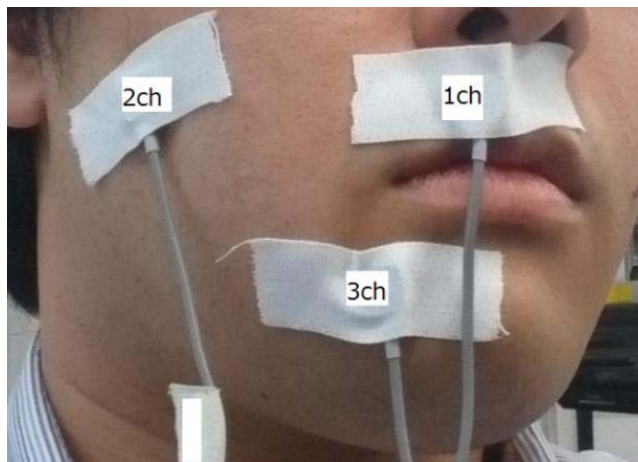


図5 電極の取り付け位置.

表1 日本語発声時における3部位の動き.

母音	下頰骨 (Mandible)	口唇 (Lips)	口角 (Angulus Oris)
あ	Pull up (上方に)	Open wide (大きく開く)	
い	Hold (維持)	Open (開く)	Pull to outside (外に)
う	Hold (維持)	Shrink/Push (縮小)	
え	Lower (下方に)	Open (開く)	Pull to outside (外に)
お	Lower (下方に)	Shrink/Push (縮小)	

本研究ではデータを分割するのではなく、1つの領域すべてを使用してFFTを適用しパワー（振幅）スペクトルの計算を行う。なお、ハミング窓で窓かけを行ってから周波数変換を行っている。

3-2-1 ノイズ処理

FFTにより得られたパワースペクトルであるが、目視で簡単に識別が可能なほど大きなノイズが多くEMG信号で確認できた。以下でそのノイズ特徴の削除及び低減手法について示す。

得られたパワースペクトルを確認すると60[Hz]の倍数帯を中心大きな成分が分布していることが確認できた。このままでは、この特徴によって本来の特徴が消え、学習が上手くいかない可能性がある。よって、

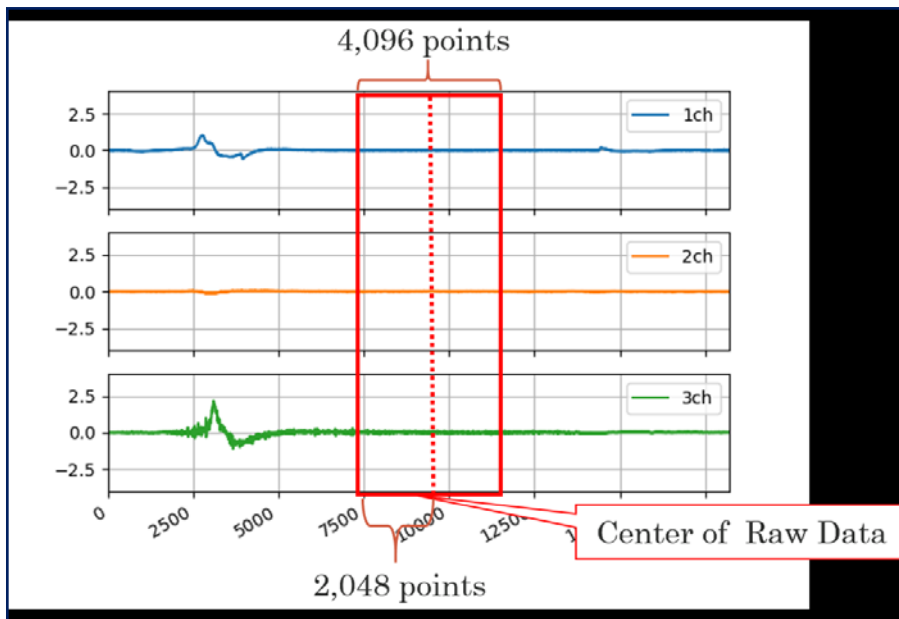


図6 基準点の決定. 長音部から切り出し.

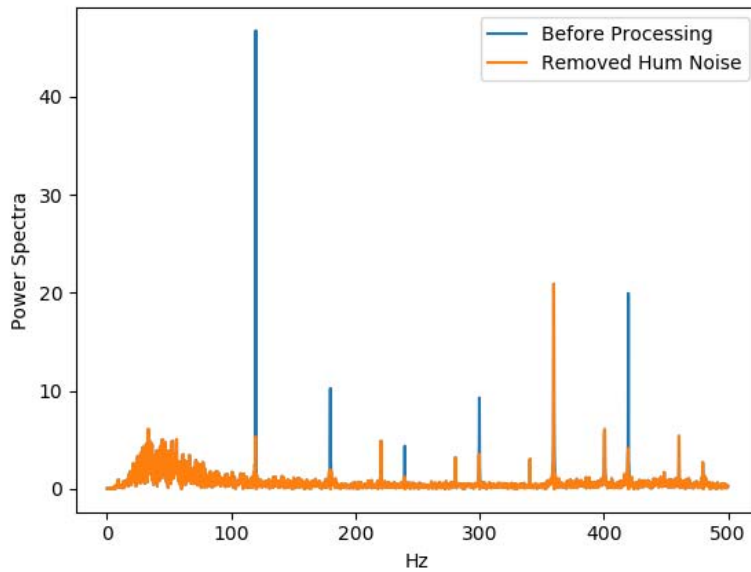


図7 ハムノイズの除去結果.

本研究ではまず、ハムノイズの削除を行う。パワースペクトルの周波数分解能の算出式は次式である。

$$\Delta F = \text{サンプリングレート} / \text{フレーム長}$$

本研究でのフレーム長は、4,096点であるため分解能 ΔF は 0.244[Hz] ということになる。本研究ではハム周波数の倍数帯に最も近い部分 F_{ref} を基準として $F_{ref} \pm \Delta F \times 1.5$ [Hz] 範囲の情報を削除した。結果を図7に示す。

次にハムノイズ以外のノイズ的特徴の低減を行う。本研究ではデータの四分位範囲の1.5倍を正常値の上下限として、以下の処理を行った。なお、 x :データ長とし、正常値の上下限はOutlierMax, OutlierMinと表記するものとする。

● $f(x) > \text{OutlierMax}$

パワースペクトル $f(x)$ が上限値を上回っていた場合、その地点を中心として、 $\Delta F \times 10[\text{Hz}]$ の範囲の値の平均値で置換を行う。

● $f(x) < \text{OutlierMin}$

パワースペクトル $f(x)$ が下限値を下回っていた場合、その値を削除し、0で置換する。

3-3 学習識別部

学習識別には深層学習手法の一つである畳み込みニューラルネットワークとSVMを使用する。なお先に述べたように、SVMを使用する際には、事前に判別分析(LDA)により次元圧縮を行うものとする。

3-3-1 データの拡張

ニューラルネットワークの学習においては、過学習が汎化性能を低下させることが指摘されている。たたみ込み層を有する深層学習ネット(CNN: Convolutional Neural Network)は過学習に頑強であるといってもMLP(Multi-Layered Perceptron)と比較した場合の話であり、過学習の影響を受けないわけではない。そのため、過学習を抑制させる意味でも学習データの水増しが行われることが多い。本研究において、データ拡張のために拡大縮小を利用する。平均1、分散0.2に従うように乱数を生成し、パワースペクトルの値を操作することで、データの総数を入力データ数 $\times 100$ 倍に拡張する。このデータ拡張により手首EMG信号認識では精度が飛躍的に改善されることを確認している[10]。

3-4 CNNのネットワーク構造

使用するCNNの構造であるが、たたみ込み層を2層適用し、その後、Maxプーリングを使用する。たたみ込み層の出力はバッチノーマライゼーションと言われる出力を平均0、分散1に標準化する処理を行う。たたみ込み層の後で、全結合の10層多層ネットワークを用いる。最終出力層は5ユニット(母音5種)となっている。たたみ込み層の構造を図8に示す。たたみ込み層の最終出力値が、次の全結合の多層ネットワーク(図9)への入力となる。各ニューロンの出力関数(非線形関数)はRectified linear unit(ReLU)を使用する。

本研究において最適化手法にはAdamを選択した。学習率は0.0001である。Adamとは勾配の平均と分散をオンライン推定し、学習率を更新する方法であり、AdaGradの勾配が疎になる場合に強い性質と、RMSPropの非定常な問題設定に強い性質の2つの長所を併せ持つ手法である[5]。学習回数は上限10,000回とし、学習が停滞して100回の間モデルの精度の向上がみられなかった場合、学習を終了することとする。モデルの評価には交差エントロピーを用いた誤差検定を用いる[5]。

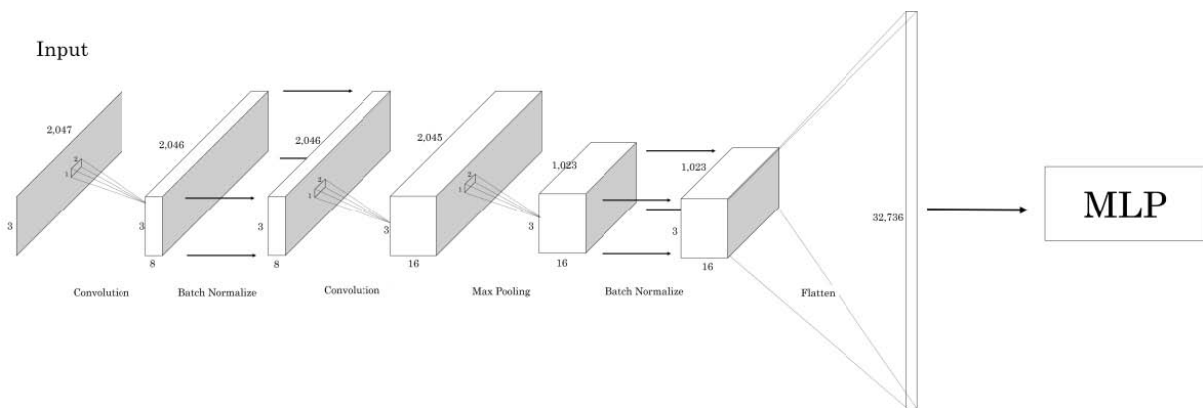


図8 たたみ込み層の構造.

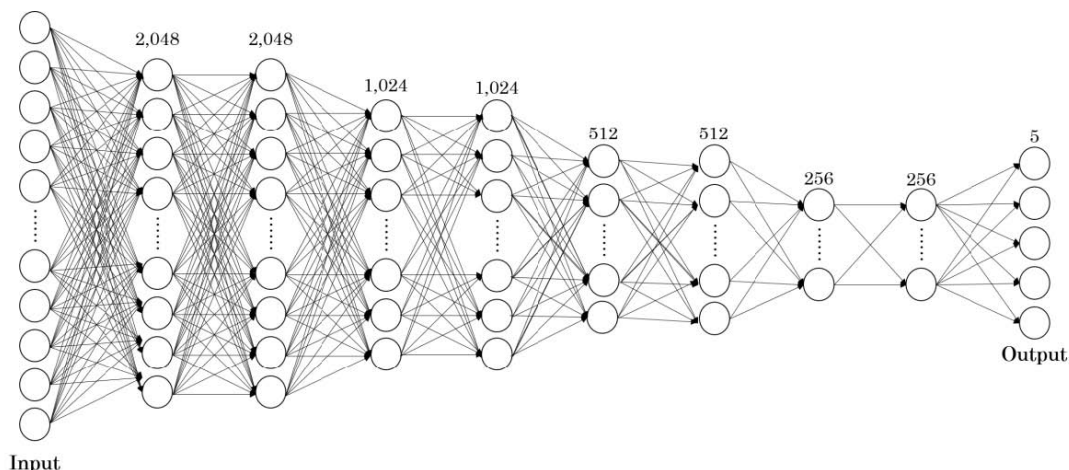


図9 全結合の多層ネットワーク (MLP).

4 母音認識の実験

本節では、まず本研究での実験条件を示す(表2)。被験者は20代男性6名 (a, b, c, d, e, f) , 1日1回の計測を3日間行った。識別対象は日本語5母音とし、被験者は各発音を10秒間行う。また、全ての発声間に同期間の弛緩状態を設ける。ただし、発声時に母音はほとんど聞こえない状態で計測する。乾式電極は非伸縮性のテーピングテープによって固定され、さらに電極の浮動を防ぐために指で押さえることで固定を行う。測定箇所は3.2で述べたように、3か所(口輪筋、大頬骨筋、口角下制筋)が測定できる場所とした(表1、図5)。

表2 測定条件.

認識対象	日本語の5母音
被験者	20代の男性6名
計測回数	3回(45データ/被験者)
センサの設置場所(筋肉)	口輪筋、大頬骨筋、口角下制筋
センサ種類	2極式乾式センサ
計測機器	P-EMG plus, サンプルング周波数1,000Hz

本研究において先行研究との結果を比較する。先行研究で性能の最高であった4,096点のデータを用いて作成したデータセットをLDA (Linear Discriminant Analysis, 判別分析) とSVMで識別した結果及び次元削減を施さずに識別した結果と比較する。CNNは同じ評価を3回行った平均値を結果とする。なおデータの条件を合わせるために、ノイズ低減手法の前処理を双方のデータに行うものとする。先行手法の識別結果は、平均で34.4%であった。提案手法の識別精度は平均で38.6%であった。結果として、約4%の改善結果が得られた。

4-1 CNNのネットワーク構造

まず、CNNにおける特徴抽出の方法は先行手法よりも4.2%向上することが明らかになった。これは畳み込みとプーリングが効果を発揮しているからであると考えられる。しかし、CNNを用いた場合でも識別率は38.6%と大きな向上には至らなかった。ここで、どの母音の誤識別の傾向を確認してみる。内訳としては「あ」と「え」(図10)、「う」と「お」の誤識別がほとんどであった。これはまず単純に口の動きがとても似通っていることに起因すると考えられる。例えば、母音「あ」を発音した時に、1chに特に顕著な特徴が現れると

考えていた。しかし、特定の被験者は3chにも特徴が現れており、奇しくも母音「お」と似通っていることが確認できた。その他にも、よく似た特徴となっているデータが観測でき、これが精度悪化の大きな要因であると思われる。

結果として「あ」「う」「お」の間と「い」「え」の間で誤識別を引き起こしたが、このように全領域にパワーが平均的にのっている、考案したノイズ低減手法では本来の成分とノイズを判断することができないため取り除くことは不可能であり、特徴抽出も難しい、個人の特徴だとするとこのままの判別は厳しいと考えられる。

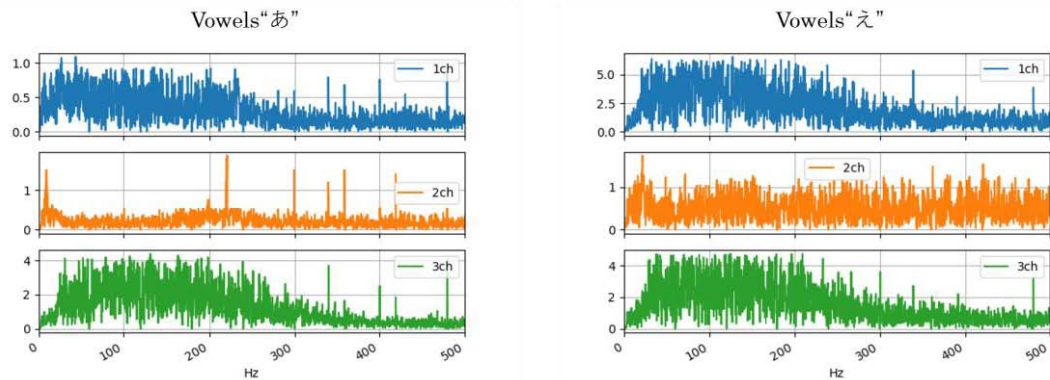


図10 被験者 e の「あ」と「え」発声時のパワースペクトル。

各被験者のパワースペクトルを確認すると基本的に信号は突発的なピークの特徴や高周波帯域においての大きな特徴を保有していることが、生信号を確認すると被験者によって信号が大きく異なることが確認できた。ノイズ的特徴に関しては、湿式電極を用いるとこのようなノイズはほとんど確認することはないため、乾式電極特有の問題であるといえる。また、外れ値除去の閾値とその検出方法、そして低減方法の再考が必要であること及びこれらのノイズを発生させにくい計測方法と実験方法の考案が必要であることが明らかになった。このようなノイズは、計測用センサの電極サイズが小さいことにも起因していると考えられる。電極サイズが大きい3極センサ(図2)を用いることも考えられるが、サイズが大きいこともあり、口唇周囲に設置することは困難である。

CNNの層構造の改良も必要であると考えられる、本研究のCNNは畳み込み2層、プーリング1層から成る単純な構造のCNNであった。しかし、これでは識別に必要な特徴が取得しきれていない可能性がある。また、畳み込み層のカーネルも他のチャンネルのことを考慮しない構造であった。したがって、フィルタサイズを変更、畳み込み層数の増加が必要であると考えられる。また、MLP部分は10層と大きな構造であったため、グローバルアベレージプーリング(GAP)に置換するといった処理も必要であると考えられる。さらに、本研究では深層学習のために乱数を用いて学習用サンプル数を増加させる方法を用いた。この乱数の使用による学習サンプル数の増加方法であるが、結果的に若干の精度改善に止まった。これは、増加させた学習サンプルのすべてが精度改善に役立ったか否かの確認をしていなかったこともその原因であると思われる。したがって今後は、発生させた学習サンプルが目的とする深層学習にどの程度役立つかの評価を行う必要がある。さらに、効率的に学習用データを発生させることも検討する予定である。

4-2 SVM と深層学習による識別実験(個人認証)

口唇EMGに基づく個人認証に関しては、時間の制約もあり、深層学習を予備的に適用しただけに終わった、そのため、ここではSVMと深層学習を用いた識別実験の結果を簡単に述べるに止める。

個人認証システムの構成は入力部、前処理部、特徴抽出部、学習識別部の4つで構成されている。入力部では筋電信号の取得を行う。前処理部では取得した信号の比較的安定した箇所を切り出す。特徴抽出部では切り取ったデータを一定数の幅でオーバーラップを設けて切り出して分割する。それらに対し窓掛け、高速フーリエ変換(FFT:Fast Fourier Transform)を適用し、振幅成分を取り出し、ノイズ除去などの処理を行う。学習識別部にはサポートベクタマシン(SVM:Support Vector Machine)と深層学習ネットワークを用いる。

SVMで使用するカーネル関数は試行錯誤の結果、線形カーネルを使用した。非線形カーネルを使用すると精度が低下したためである。

使用した電極は母音印式と同じ小型の2極乾式電極であり、口唇周囲への設置場所も同じである。EMG信号の計測を行う被験者数（6名：a, b, c, d, e, f）と前処理方法は母音認識と同様である。

Leave-one-out クロスバリデーションを行うことにより、個人認証の結果として、「あ」～「お」のデータに対して平均で、ノイズ除去を行わない場合（Data 1）で55.1%、ノイズ除去を行った場合（Data 2）に58.6%の精度を得た。Data 2で、母音の「お」、被験者Cにおいてデータセット1では1.06%、データセット2では8.99%、データセット3では100.0%というように、個人内差が目立った。原因としては、乾式電極の使用が考えられる。測定時に電極が浮く、位置が若干ずれるなどの状況が発生していたため、これによって生じたアーチファクトなどによって周波数領域にノイズが混入した可能性がある。よって測定時の乾式電極の設置方法を改善する必要がある。現在の商用周波数の倍数帯周辺のノイズ除去は、前後を含めて5点の平均を取ったものであるが、ノイズを低減しきれていないデータも見受けられた。これは外れ値が極端に大きいと平均をとったとしても外れ値に引っ張られ、十分に低減することができないためと考えられる。したがって、よりの確に除去ができるようなノイズ除去方法の導入が必要である。例えば、大きく外れている値は最初から排除しておく、などの措置は必要である。

次に、たたみ込み層を有する深層学習ネットワークを用いた予備的な実装結果について述べる。4層のたたみ込み層と3層の全結合ネットワークを組み合わせて個人認証の予備実験を行った。その結果、「あ」～「お」の5母音に対する個人認証の結果として平均で69.6%の結果が得られた。この結果は、SVMの結果よりも優れており、深層学習の有用性が実証できた。しかし、この結果はあくまでも予備的な実験結果であり、他のネットワーク結合や、十分な試行回数を試すことができず、またプーリング層を含めた実験も行っていない。これらは、今後の課題である。

以上のように、本研究での学習識別は主にSVMでの実装であったが、今後は機械学習、特に畳み込みニューラルネットワークなどの深層学習の導入を強力に進めていく予定である。この際、個人認証と母音認識を同時に一つの深層学習ニューラルネットワークで実現することを検討している。個人認証と母音認識の同時実現は一見すると相反する目的を同時に実現することのように見える。個人認証は個人の特徴を検出する必要があるが、母音認識は各個人の特徴には依存しない状況で各母音を認識する必要があるためである。しかし、過去には文献[11]で示されているように、回転不変性と回転依存性の相反する性質を一つの3層ニューラルネットワークで実現した例が報告されている。このニューラルネットワークでは、最初から二つの相反する性質を同時に実現するための制約がネットワーク重みに課されていた。学習中も二つの相反する性質を同時に実現する仕組みが保存され、最終的に高精度の回転不変パターン認識と回転角度推定を実現できている。深層学習では、多数の学習サンプルを準備することにより高精度認識が実現できている。したがって、多数の学習サンプルを用いる深層学習で二つの相反する性質である個人認証と母音認識を同時に実現すること目指し、またそのためのネットワーク構造を検討していく予定である。たたみ込み層のネットワーク部分で前述の個人認証と母音認識の特徴をうまく学習することができれば、両方の精度を改善できる可能性がある。今後は、この方法について考察を行いながらシミュレーションを実装していく予定である。

5 おわりに

本研究では、口唇周囲のEMG信号（表面筋電信号）に着目してサイレント音声認識と個人認証を行った。通常のサイレント音声認識では母音認識は比較的に高精度に認識可能であるが子音認識は精度が低い状態が続いているためである。また、従来の研究で使用するEMG計測センサは湿式が多かったが、実用性を考慮して乾式センサを使用して、まず母音認識での高精度化と個人認証を目指した。

日本語5母音の認識と個人認証を行う場合に、EMG信号の計測時にノイズが重畳することが多く、まずはその低減化を行った。ノイズの低減化を行った場合と行っていない場合を比較して実験を行った。サイレント音声認識（5母音認識）では、深層学習を用いる提案手法の識別精度は平均で38.6%であった。従来法よりも、約4%の改善に止まった。このように精度は比較的低い状態となっている原因は、計測時に様々なノイズが重畳していることと計測データ数が少ないことが考えられる。今後はノイズの特性をさらに把握した上で、ノイズ低減下方法の検討および、学習に用いるデータ数を増加させる方法を考案していく必要がある。学習に用いるデータ数を増加させる方法は、申請者のグループでも提案し、精度改善に効果があることを確認

している。しかし、増加させたデータが深層学習での学習に対して真に効果的なデータとなっているか否かの評価は行われていない。今後は、この追加データを評価する方法を開発することが重要である。

次に、個人認証であるが、この場合もノイズを低減した場合としていない場合で比較した。その結果、SVMを用いた場合に平均精度は58.6%であり、3%程度の改善に止まった。個人認証では識別器にSVMを主に使用したが、現時点で高精度を達成できていない。一方、予備的な実験として深層学習を導入したが、時間の制約で十分な評価が行えていない。ただ、現時点で、約69.6%の精度を達成できている。今後は深層学習の導入を積極的に進め、高精度認証を目指す必要がある。

以上のように、本研究では母音認識と個人認証で高精度化を目指したが、乾式電極を用いた場合に、精度は湿式電極を用いる場合に比べて低い状態に止まった。今後は、乾式電極の特性を詳しく把握した上で、さらにノイズの影響を低減化する方法の研究を実施予定である。また、深層学習ニューラルネットワークの構造をサイレント音声認識と個人認証に適した構造に改良することと、学習に必要なデータ数を効率的に増加させる方法の検討および増やした学習データの有効性の評価が重要であると考えられる。

【参考文献】

- [1] 坂倉淳, “無喉頭音声習得方法と発声機構人工喉頭音声”, 音声言語医学39.4, pp. 443-449, (1998)
- [2] 高橋毅, 景山陽一, 西田眞, 若狭亜希奈, “口唇の色彩情報および形状情報に着目した発話フレーム検出法”, 知能と情報23.2, pp.146-156, (2011)
- [3] 永井秀利, 中村貞吾, 野村浩郷, “自然言語インターフェースのための無発声音声認識への活用を目的とした表面筋電波形の分析”, 信学会技術報告TL, 思考と言語102.688, pp.25-32, (2003):
- [4] 木塚朝博, 木竜徹, 増田正, 佐渡山亜兵, “表面筋電図”, 東京電機大学出版局, (2006)
- [5] 人工知能学会連載解説「Deep Learning (深層学習)」(第1回～7回), 人工知能学会誌, 28巻3号～第29巻4号 (2013年～2014年)
- [6] K. Fukushima, Neocognitron, “A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position”, Biological Cybernetics, Vol. 36, No. 4, pp. 193-202, (1980)
- [7] Y. LeCun et al. “Gradient-Based Learning Applied to Document Recognition”, Proceedings of the IEEE, 86(11): pp.2278-2324, (1998)
- [8] 追坂電子機器, <http://www.oisaka.co.jp/p-emgplus.html>
- [9] Ryosuke Takabatake, Shin-ichi Ito, Momoyo Ito and Minoru Fukumi, “Japanese Vowels Recognition Using Linear Discriminant Analysis and Surface Electromyogram Measured with Bipolar Dry Type Sensors”, Proceedings of the 5th IIAE International Conference on Intelligent Systems and Image Processing 2017, Waikiki, Hawaii, USA, (2017)
- [10] Daiki Hiraoka, Shin-ichi Ito, Momoyo Ito and Minoru Fukumi, “Japanese Janken Recognition Based on Wrist EMG Analysis by CNN and SVM”, Proceeding of International Conference on Advanced Technology & Sciences (ICAT'Rome), pp.323-328, Rome (2016)
- [11] Minoru Fukumi, Sigeru Omatu and Yoshikazu Nishikawa, “Rotation-Invariant Neural Pattern Recognition System Estimating a Rotation Angle”, IEEE Transactions on Neural Networks, Vol.8, No.3, pp.568-581, (1997)

〈発表資料〉

題名	掲載誌・学会名等	発表年月
乾式電極を用いた口唇 EMG による個人認証	電気学会・電子情報システム部門大会	2018年9月