

## 科学研究費助成事業 研究成果報告書

平成 29 年 6 月 13 日現在

機関番号：17104

研究種目：基盤研究(C) (一般)

研究期間：2013～2016

課題番号：25330198

研究課題名(和文) 表面筋電信号に基づくウェーブレット係数の重心推移法を活用した黙声認識に関する研究

研究課題名(英文) Research on Inaudible Speech Recognition based on the Center-of-Balance Transition Method on Wavelet Coefficients of Surface EMG Signals

研究代表者

永井 秀利 (NAGAI, Hidetoshi)

九州工業大学・大学院情報工学研究院・助教

研究者番号：60237485

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：声を出さずに自然に発声された内容を口唇周辺の表面筋電を用いて認識することを目指し、ウェーブレット係数の重心推移法を活用した認識手法の研究を行った。自然な連続発声は口唇形状を緩やかに変化させつつ行われるため、認識のための特徴を抽出すべき位置を同定しなければ、良い認識結果を得ることは難しい。そこで、重心推移を活用して発声変化時点や口唇形状が最も明瞭となった時点を同定する手法を提案した。しかしながら、連続母音認識結果の1位正解率は60%程度であり、更なる改良が必要と言えた。

研究成果の概要(英文)：The purpose of this research is to recognize natural inaudible speech based on surface EMG signals around lips. During this research period, I studied the recognition method using the center-of-balance transition method on wavelet coefficients of surface EMGs. A person performs a natural continuous utterance with gentle change of lip shape. Therefore, in order to obtain a good recognition result, it is necessary to identify the proper positions to extract the features for recognition. As such positions, I propose methods to detect the change point of phonemes and the peak point of lip shape change based on center-of-balance transition. However, the first place accuracy rate of the recognition result of continuous vowels is about 60%. It shows that further improvement is necessary.

研究分野：生体信号処理

キーワード：黙声認識 無発声音声認識 サイレント音声認識 表面筋電 ウェーブレット解析 ウェーブレット重心 重心推移法

### 1. 研究開始当初の背景

音声認識は有用な技術ではあるが、騒音下やプライバシー確保などの理由で声を出せない状況での利用は難しく、声帯摘出などで発声ができない場合には全く利用できない。そのような状況で音声認識と同等の応用を可能にする技術として、一般的なマイクで計測できるレベルの可聴音を用いずに発声内容を認識する技術が求められていた。

可聴音を用いない発声の認識は、マイクで拾えないようなささやき音声を用いるNAM(Non-Audible Murmur)認識技術や、音声信号を一切用いずにカメラ画像や表面筋電に基づく認識技術の研究が行われている。表面筋電による認識をカメラ画像による読唇と比較した場合、短所は表面筋電計測のための電極装着が必要なことであり、長所はカメラ撮影が困難な状況(マスク装着、暗闇、激しい照明変化など)でも利用可能であること、および、アクセントのような強弱情報を計測できる(例えば「橋」と「箸」とを文脈によらずに判別しうる)ことである。

研究開始当初において、表面筋電を用いた認識は、訓練などにより口の形を強調するようにした母音の認識は十分な精度を得ることができていた。口形を強調した母音であれば、認識精度を高めることは容易である。しかしながら、口形を強調して話し続けることは負担が大きく、さらには、子音を口形強調して発声することは極めて困難である。また、口形強調を前提とした手法を単純に拡張して自然な発声に対応させるというのは実現の見込みに乏しい。よって、自然な発声を前提とした認識手法が求められていたが、孤立母音であれば95%程度の認識率を達成できたものの、連続発声や子音については不十分な手つかずに近い状況と言えた。

### 2. 研究の目的

本研究の最終目標は、口唇周辺から頸部の表面筋電を利用して、声を出さずとも発声時と同様に自然に口を動かすだけでその内容を認識する(黙声認識と呼ぶ)の実現である。

本研究での計測対象となるのは、主に表情筋である。骨格筋と比較した場合、表情筋は小さい筋であるために信号が弱い。また、発声に伴って短時間で急激に変化する活動を的確に捉える必要がある。表面筋電に基づいて筋活動を捉える際に用いられる一般的な手法はRMSや平均パワー周波数などであるが、そうした一般的な手法は微弱な活動の検出能力や、変化を捉える際の時間解像度が不足する。そこで研究代表者は、一般常識的に必要十分と言われる表面筋電の計測周波数よりも遥かに高い周波数で計測した信号を用いたウェーブレット解析結果に基づいて、高い時間解像度で筋活動変化を追跡する重心推移法(Center-of-Balance Transition Method)を提案し、その有効性を示した。

その結果を踏まえ、最終目標に向けての

ステップとして、本研究期間には黙声子音特徴の抽出や連続黙声認識を行う技術の構築を目標とした。また、現状では任意の発声を十分な精度で認識することはできないにしても、語彙が限られていれば十分な認識精度が得られる可能性がある。そこで、提案する技術を応用して少数語彙世界での単語認識も試みる。

### 3. 研究の方法

口唇形状の強調はせず、自然な発声を対象とするという点は、本研究の根幹とも言える重要な立ち位置である。この場合、母音ですら発声期間中の特徴の定常性を期待できないという大きな問題が存在する。音声認識等において、母音は発声期間中の特徴の定常性が高いものとして捉えられ、認識にも活用される。この定常性は、口唇形状を強調する場合であれば期待することもできるだろう。しかし、自然な発声は口唇形状を緩やかに変化させつつ行われるため、長音発声でない限りは母音であっても筋活動が定常的な時間区間は無きに等しい。そのため、任意に切り出した時間区間に対して、その区間内の特徴のみから認識を行おうとしても、発声間の推移途中の曖昧な状態が多数となり、良質な判別が可能となるような特徴は得られない。

また、子音の発声に伴う口唇変化は母音に比べると急速であるために短時間での活動変化を追跡する必要があるが、そのように短い時間区間で信号特徴を切り出して認識しようとした場合、一時的なノイズや母音発声途中の活動の揺らぎを誤って捉えてしまう危険性が高い。

これらの点を鑑みると、発声に伴う活動推移の大きな流れの中から認識のための基準となる位置を検出し、その位置を基準とした領域から特徴抽出することが効果的である。そこで本研究課題では、連続黙声認識のための基準位置として、連続する2音の間で発声が切り替わる位置(発声変化位置)と、母音発声時に口形が最も強く形作られた時点(口形極点と呼ぶ)とを設定し、その位置を推定する手法の開発を目指した。

子音の特徴抽出に関して、本研究では音素としての/k/や/m/ではなく、後続する母音を含めた/ka/や/mi/といった単位での抽出を考える(単純に「子音」と呼ぶときには、この単位を指すものとする)。これは、後続する母音に依存して口唇の動きが変化するためである。

子音の発声特徴は発声変化位置直後に現れると考えられる。そこで、発声変化位置に基づく連続黙声認識において、母音と一部子音との2連続音を対象とした実験を試みる。

子音の内、両唇音については口唇に大きな動きを伴うため、比較的単純な筋活動量の大小変化を捉えることで認識できる見込みは大きい。しかし、両唇音以外の子音は、両唇音ほどには口唇の顕著な動きを伴わないた

め、認識のための特徴を捉えることは難しい。そこで本研究課題では、孤立子音のデータを収集し、単純な筋活動量の大小に限らない特徴獲得を目指して分析を行った。

#### 4. 研究成果

##### (1) 発声変化位置に基づく連続黙声認識

2連続音の表面筋電波形のウェーブレット解析結果の時間-周波数平面を分析すると、先行音発声のための口唇形状形成を完了した後、後続音発声のための口唇形状形成に移り変わる際に一時的な筋活動の低下が観測される(図1)。この活動低下は、重心推移においてウェーブレット重心値の一時的な低下として現れる。この低下を発声変化位置として検出する。

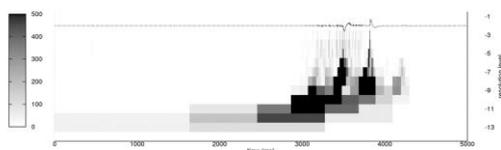


図1 /ウオ/発声時の下唇下制筋の表面筋電のウェーブレット解析結果

従来は重心推移の局所的な谷の深さによって検出を目指していたが、その手法は、発声速度の変化や閾値の設定への依存性が高いという欠点があった。そこでウェーブレット重心の差分推移と呼ぶ手法を提案し、筋活動が減少から上昇に切り替わる位置をより明瞭にした。しかしながら、そのような活動変化の時間位置はすべての計測チャンネルで同一ではない。そこで、各チャンネルの差分推移を集約し、発声変化位置を同定する手法を提案した。

/ア/, /カ/, /サ/, /マ/, /ワ/, /ン/の内の2音の組30種の連続発声データを用い、提案手法による発声変化位置の推定と、発声変化位置付近の窓区間のウェーブレット重心値をパラメータとした認識実験を行った。発声変化位置の再現率は66.6%、適合率は69.2%であった。2音の組としての認識精度は、正しい発声変化位置を与えた場合の1位正解率66.3%、5位正解率98.3%であり、変化位置検出から認識までを通して実行した場合の1位正解率51.0%、5位正解率89.3%であった。

##### (2) 口形極点に基づく連続黙声母音認識

母音の特徴は口唇形状に現れやすいが、自然な母音の連続発声での途中状態は必ずしも明瞭な特徴を有していない。とはいえ、口唇形状の変化の流れにおいては、口唇形状が最も強く形作られる瞬間が存在するはずである。そのような瞬間を口形極点と呼ぶ。口形極点を検出し、その時点の特徴を認識パラメータとして用いるようにすれば、認識精度の向上を期待できる。

口形極点は、口唇の開きの縦幅が局所最大、横幅が局所最大ないし最小となる時点と考

える。そのため、口唇周辺の表情筋の発揮力評価に基づいて口唇形状変化を推定、追跡し、口形極点の位置を推定することとした。

表面筋電に基づく一般的な筋の発揮力評価では信号のRMS値を用いるが、RMSでは表情筋の弱くて変化の速い表面筋電信号の活動を十分には捉えることができない。また、RMS値は筋疲労によって特性が大きく変化することも良く知られている。そこで、冗長ウェーブレット解析に基づく筋活動特徴量の体系化を提案し、その特徴量を口唇形状変化の推定に用いることとした。ただし、体系化では筋疲労影響を加味した発揮力評価を提案しているが、今回の研究で用いるデータは時間が短く、筋疲労の影響は乏しいと考えられるため、疲労に関する特徴量を省いて簡易的に、周波数特徴量であるウェーブレット重心と信号強度特徴量であるウェーブレットRMSとの積で評価した。

この積の値を各筋の表面筋電から求め、縦横それぞれに各筋の影響力を鑑みた重みづけをして加算し、各幅の変化を推定した。縦横それぞれの極点から口形極点位置を推定するが、それぞれの極点位置は必ずしも一致せず、また、筋活動の揺らぎなどで口形極点候補は複数存在する。そこでいくつかのルールによって絞り込んだ後、最終的には1音1点ずつのすべての組み合わせの中で確率最大となるものを口形極点推定および認識の結果とした。

連続2母音のデータを用いた認識実験では、口形極点検出における再現率が92.6%、適合率が80.5%であった。また、口形極点検出から認識までを通して行った場合の2連続音同時正解率は60.5%であった。

##### (3) 少数語彙世界における黙声認識

介護時の要請を想定し、「手、洗う」のような2単語文(先行語5語、後続語4語)と「飲む」、「食べる」、「はい」などの1単語文(6語)からなる少数語彙世界を設定した。入力は文単位とし、2語文の場合は単語を区切りつつ発声する。この単語間のギャップによって1語文か2語文かは高い精度で分類可能であり、切り出した単語発声区間がどの種類の単語かも同定できる。

各単語の認識手法は発声変化位置に基づく手法に準じることとしたが、語長が異なるために検出すべき変化位置の数をあらかじめ決めておくことはできない。そのため、多数発生しうる発声変化位置の候補の適切な絞り込みは難しい。しかし、何個の発声変化位置が存在すると捉えられやすいかには、単語ごとに傾向の違いがあると予想できる。これを確率で扱い、認識パラメータの一つとして組み込んだ。ただし、この確率を統計的に得るのに十分なサンプルを収集することは困難であるため、変化位置候補検出のための閾値を確率的に動かすモデルを作成し、変化位置検出数の確率推定値を与えた。

認識実験での1位正解率は、1語文の単語で89.4%、2語文先行語で51.0%、2語文後続語で80.0%であった。

#### (4) 孤立子音の特徴分析

子音の発声には舌の動きが重要であるが、舌の形状や動きを左右する舌筋などは表面筋電では計測することができない。そのため、口唇周辺の筋の協調動作のわずかな違いなどから推定せざるを得ず、子音の認識は極めて難しい。機械学習を活用するにも、学習がうまく機能するだけのサンプルを集めることが困難である。そこで、日本語単音の特徴の類似性でいくつかのグループに分類した上でグループごとに特有の手法や言語処理の適用により認識を行うことを目指し、促音、拗音を除く日本語単音70種のデータを収集して分析を行った。

分析の結果、各チャンネルの重心推移がピークを迎える順序関係に特徴差が見受けられたため(図2)、この特徴を用いた日本語単音のクラスタリングを試みた。

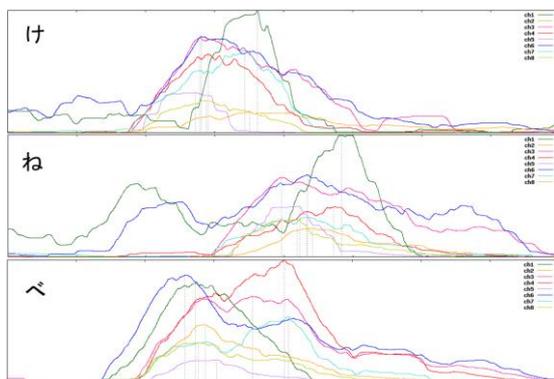


図2 チャンネルごとの筋活動ピーク位置出現順の差異

今回は、各チャンネル組の前後関係のみを単純に真偽値で表したものを特徴ベクトルとして用いた。結果は、「少しではあるが有効性が認められる」という程度であったが、すべてのチャンネル組を用いるよりも基準チャンネルを限定した組のみを用いる方が良好という知見を得た。これは、ある発声に寄与する筋は計測対象筋のすべてではないことが原因と考える。認識等において、寄与していない筋のパラメータも同等の重みで評価されることで、寄与割合の高い筋の評価の重要性が低下しているためと推測する。この問題は連続母音認識などの他の認識においても同様に存在するものであり、今後の黙声認識においてこの寄与率問題をどのように組み込んでいくかが重要な課題と言える。

#### 5. 主な発表論文等

[雑誌論文] (計 11 件)

① 永井 秀利: A Method for Real-Time Estimation of Local Muscular Fatigue in Exercise Using Redundant Discrete Wavelet Coefficients, 3rd

International Conference on Systems and Informatics (ICSIAI 2016), 査読有, 2016, pp. 819~825, DOI: 10.1109/ICSIAI.2016.7811064

② 黒木 菜緒, 永井 秀利, 中村 貞吾: Continuous Inaudible Recognition of Japanese Vowels Using Features Detected at Lips Shape Peaks based on Surface Electromyography, 3rd International Conference on Systems and Informatics (ICSIAI 2016), 査読有, 2016, pp. 959~964, DOI: 10.1109/ICSIAI.2016.7811089

③ 永井 秀利: 表面筋電の冗長ウェーブレット係数に基づく筋活動の量的および質的特徴量の体系化, 電子情報通信学会技術研究報告, Vol. 115, No. 513, 査読無, 2016, pp. 119~124

④ 黒木 菜緒, 永井 秀利, 中村 貞吾: Continuous Inaudible Vowels Recognition based on Features of Surface EMG at Candidates for Peaks of Lip Shape, Proceedings of Life Engineering Symposium 2015, 査読有, 2015, pp. 325~328

⑤ 永井 秀利: Computational Complexity in Continuous Estimation of Muscular Activity based on Redundant Wavelet Coefficients of Surface EMG, Proceedings of Life Engineering Symposium 2015, 査読有, 2015, pp. 329~334

⑥ 永井 秀利: A Method to Extract Minute Characteristics of Surface EMG using Redundant Wavelet Coefficients, Proceedings of the 3rd International Symposium on Advanced and Applied Convergence (ISAAC 2015), 査読有, 2015, pp. 138~143

⑦ 永井 秀利: 冗長ウェーブレット解析に基づく運動中の筋疲労の実時間評価手法, 電子情報通信学会技術研究報告, Vol. 115, No. 383, 査読無, 2015, pp. 37~42

⑧ 永井 秀利: 冗長ウェーブレット解析に基づく重心推移法による筋活動量評価手法, 電子情報通信学会技術研究報告, Vol. 115, No. 49, 査読無, 2015, pp. 15~20

⑨ 黒木 菜緒, 永井 秀利, 中村 貞吾: 表面筋電より推定する口形極点を活用した連続黙声母音認識, 電子情報通信学

会技術研究報告, Vol. 115, No. 36, 査読無, 2015, pp. 243~248

- ⑩ 永井 秀利, 金井田 早紀, 中村 貞吾: ウェーブレット重心の推移特徴に基づく黙声変化位置検出と連続黙声認識での活用, 電子情報通信学会技術研究報告, Vol. 115, No. 36, 査読無, 2015, pp. 249~254
- ⑪ 金井田 早紀, 永井 秀利, 中村 貞吾: 表面筋電信号に基づく黙声子音の発声変化位置候補検出のための一手法, 電子情報通信学会技術研究報告, Vol. 114, No. 79, 査読無, 2014, pp. 45~50

[学会発表] (計 10 件)

- ① 黒木 菜緒, 永井 秀利, 中村 貞吾: 連続音間の相関パラメータを導入した口形極点に基づく連続黙声母音認識, 第3回サイレント音声認識ワークショップ, 2016/10/14 ~ 2016/10/15, 福岡朝日ビル貸会議室 16号室会議室 (福岡県福岡市)
- ② 永井 秀利, 渡邊 優実, 中村貞吾: 少数語彙世界の特性を活用した黙声音声認識手法に関する一提案, 第3回サイレント音声認識ワークショップ, 2016/10/14 ~ 2016/10/15, 福岡朝日ビル貸会議室 16号室会議室 (福岡県福岡市)
- ③ 黒木 菜緒, 永井 秀利, 中村 貞吾: 口形極点に基づく連続黙声母音認識への連続音間での相関パラメータ導入の効果, 第69回電気・情報関係学会九州支部連合大会, 2016/09/29 ~ 2016/09/30, 宮崎大学 木花キャンパス (宮崎県宮崎市)
- ④ 永井 秀利: 冗長ウェーブレット解析に基づく微細な筋活動特徴の抽出手法, 第2回サイレント音声認識ワークショップ, 2015/10/16~2015/10/17, 神戸大学 瀧川記念学術交流会館大会議室 (兵庫県神戸市)
- ⑤ 永井 秀利, 金井田 早紀, 中村 貞吾: ウェーブレット重心の推移特徴による黙声変化位置検出を用いた連続黙声認識の一手法, 第2回サイレント音声認識ワークショップ, 2015/10/16 ~ 2015/10/17, 神戸大学 瀧川記念学術交流会館大会議室 (兵庫県神戸市)
- ⑥ 黒木 菜緒, 永井 秀利, 中村 貞吾: 口形極点を特徴抽出基準位置とした連続黙声母音認識, 第2回サイレント音声

認識ワークショップ, 2015/10/16 ~ 2015/10/17, 神戸大学 瀧川記念学術交流会館大会議室 (兵庫県神戸市)

- ⑦ 黒木 菜緒, 中村 貞吾, 永井 秀利: 連続黙声母音認識における口形極点候補の組み合わせ数削減, 第68回電気・情報関係学会九州支部連合大会, 2015/09/26~2015/09/27, 福岡大学 (福岡県福岡市)
- ⑧ 金井田 早紀, 永井 秀利, 中村 貞吾: 筋活動の増減傾向に基づく子音を含む連続黙声の発声変化位置候補検出, 第67回電気・情報関係学会九州支部連合大会, 2014/09/18~2014/09/19, 鹿児島大学 郡元キャンパス (鹿児島県鹿児島市)
- ⑨ 金井田 早紀, 永井 秀利, 中村 貞吾: 筋活動変化の時間差に基づく連続黙声中の両唇音の特徴の分析, 第66回電気・情報関係学会九州支部連合大会, 2013/09/24~2013/09/25, 熊本大学 黒髪南地区 (熊本県熊本市)
- ⑩ 金井田 早紀, 永井 秀利, 中村 貞吾: ウェーブレット係数の重心値の差分推移に基づく黙声両唇音の特徴分析, FIT2013, 2013/09/04~2015/09/06, 鳥取大学 鳥取キャンパス (鳥取県鳥取市)

## 6. 研究組織

### (1) 研究代表者

永井 秀利 (NAGAI, Hidetoshi)  
九州工業大学・大学院情報工学研究院・助教  
研究者番号: 60237485