

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 5 月 28 日現在

機関番号：17104

研究種目：基盤研究(C)

研究期間：2010 ～ 2012

課題番号：22500157

研究課題名（和文） 口唇周辺から頸部にかけての表面筋電に基づく黙声認識に関する研究

研究課題名（英文） Research on Inaudible Speech Recognition Based on Surface EMG at the Area Around Lips and Neck

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

研究成果の概要（和文）：

声を一切出さずに発声した内容（いわゆる「ロパク」）を、表面筋電信号に基づいて認識することが本研究の目的である。微弱でなめらかに推移する自然な発声時の表面筋電信号の特徴を的確に捉えるために、ウェーブレット係数の重心推移法（Center-of-Balance Transition; CoBT）と呼ぶ手法を提案した。さらに、重心推移法を活用して、無理に口唇を形作らない自然な発声でなされた連続黙声 2 母音の発声変化位置を 90% 程度の精度で同定することに成功した。

研究成果の概要（英文）：

The purpose of this research is recognition of inaudible speech (move the lips without speaking) based on surface EMG signals. We proposed the method which is called “Center-of-Balance Transition of wavelet coefficients (CoBT)”. It can treat features of extremely weak and smoothly changing sEMG on natural inaudible speech. And as an application of CoBT, we tried to detect boundaries of natural continuous inaudible vowels, and acquired about 90% of accuracy.

交付決定額

（金額単位：円）

	直接経費	間接経費	合 計
2010 年度	3,100,000	930,000	4,030,000
2011 年度	400,000	120,000	520,000
2012 年度	200,000	60,000	260,000
年度	0	0	0
年度	0	0	0
総 計	3,700,000	1,110,000	4,810,000

研究分野：総合領域

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：黙声認識、無発声音声認識、表面筋電、ウェーブレット解析

1. 研究開始当初の背景

完全な無発声ではない微発声については NAM (Non-Audible Murmur) 認識の研究も存在するが、本研究は一切の発声が必要ない（発声のための振動源が必要ない）こと、またそれゆえに呼吸のリズムとは無関係に音声入力可能であることから、より広い範囲での

活用が可能である（図 1）。例えば、声帯を失った人や酸素マスク等で呼吸補助を受けている状況の人でも最低限の負担で音声入力やコミュニケーションを可能にできる。

呼吸吐出すらほとんどないような完全な無発声を前提とした認識を目指す本研究は、世界的に見ても極めて独創的なものである。

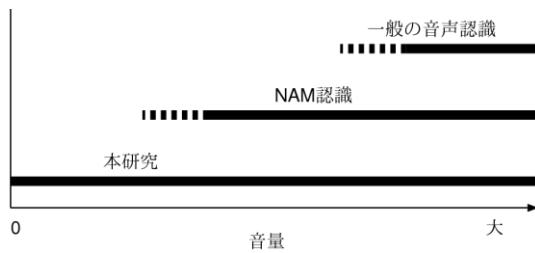


図 1. 発声内容の認識に関連する各研究分野がカバーする音量の範囲

研究開始当時、英語については米国を中心に研究活発化の動きが見られ、日本語においてもいくつかの研究発表もなされてきており、一つの研究分野として花開きつつあると言える状態にあった。

言語に依存した特性の違いは大きいので、どの言語を対象にするかは重要な要素である。本研究は日本語を対象とするため、特に日本語を対象とした他の研究と比較する。少なくとも当時の他の研究では、母音を対象として口唇形状を無理に形作るように訓練することで認識率を稼ぐ方法を取っていた。しかしながら、そのような方法は現在の認識率を稼ぐことはできても、長時間連続しての発声には苦痛を伴うために実用に耐えない。さらに、そのような方針での研究の延長線上では、苦痛の少ない自然な発声の子音を含めて認識することはほぼ不可能であるというのが研究代表者の主張であった。

2. 研究の目的

本研究全体の目的は、発声時と同様に口を動かすだけで、声を一切出さずとも音声認識と同様の計算機への音声入力・操作(「黙声認識」と呼ぶ)を可能にすることにある。本研究は、音声認識に比べて耐騒音性や情報秘匿性に優れるだけでなく、黙声/発声の切替えによる計算機入力と会話とのシームレスな切替えや声帯摘出者の発声代行など、音声認識だけでは困難なことの実現にもつながる。

研究代表者による従来の研究では、自然な発声での黙声単母音に対して、発声開始時の複数の時区間の表面筋電信号強度に基づく特徴量を用いることで 95%を超える認識率を獲得することができた。しかしながら、設定する時区間には、話者の発声リズムに依存した最適の区間幅と言えるものが存在し、そこから遠ざかるにつれて、正負それぞれに認識率が単調に減少することが示された。よって本研究課題では、話者の発声リズムを抽出して区間幅に反映する手法、あるいは発声リズムの変化に対して耐性を持つような手法の開発を目指した。

黙声母音に比べ、黙声子音の認識は大幅に難しくなる。母音は定常性が強く、「状態」で判別できる可能性が高いが、子音は刹那的

で、瞬間的な「推移」を適切に捉える必要がある。しかしながら、本研究課題に至るまでの研究により、単純に時区間を短く取のでは黙声母音の認識率を大幅に低下させてしまうことがわかっていった。そこで本研究課題では、黙声子音の発声特徴を得るのに有効な筋電計測位置を迫及すると同時に、時区間の幅はある程度を維持しつつ、「推移」うまく表せるようなパラメータ化手法を開発し、黙声子音認識の基礎を築くことも目標とした。

3. 研究の方法

黙声認識のポイントとなるのは、表面筋電信号をどの位置で何か所計測し、計測された波形データをどのようにパラメータ化するかである。通常の音声認識であれば、入力は 1 チャンネルの音声信号と確定していると言えるが、筋電を用いる場合にはどのような入力(計測対象筋)集合とすべきかについてすら確定してはいない。また計測対象筋の集合を定めたとしても、計測のための電極をどの位置に装着するかによって、その有効性も大きく変化する。そのため本課題の研究の主題は、大きく分けて「筋電計測位置に関する研究」と「認識手法に関する研究」とを主、「応用システムに関する研究」を従として行うこととして、研究を開始した。

しかしながら、認識手法に関する研究を進める上で表面筋電波形データの分析を行っている際に、ブレイクスルーとなるような極めて重要な知見を得たため、以降はこの知見とその活用とに研究の中心を移した。他の主題については、この新たな知見に基づいて見直しを行い、補助的に研究を進めた。

4. 研究成果

表面筋電信号に基づいて筋の活動量を評価する方法は複数存在するが、通常的手法はいずれも信号強度(信号の積分量)に依存したものである。本課題以前に研究代表者を取り組んだ黙声単母音認識における特徴量化の手法もその一つと言える。当初はその延長での特徴量化を目指して研究を行ったが、黙声単母音ならばともかく、自然な連続黙声や黙声子音の認識を考えた場合、信号強度に基づく特徴量だけを用いていたのでは、的確に特徴を捉えることは不可能に近いとの結論に至った。

図 2 は、[イイ]と連続で発声した際の口唇周り 4 か所の表面筋電信号にノイズ低減処理を加えた後の波形の例である。このデータを眺めた場合、チャンネル 3 には強い活動が観測されるが、チャンネル 1 には活動がなく、チャンネル 2、4 についてもほぼ活動なしと判断するのが普通であろう。また、信号強度のみで二つの[イ]の境界を同定するのは困難であることも予想できよう。

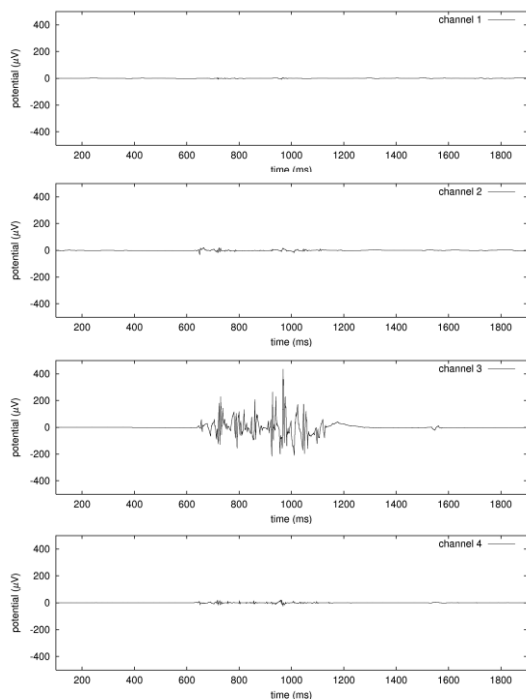


図 2. ノイズ低減処理後の連続母音[イイ]発声時の表面筋電波形の例

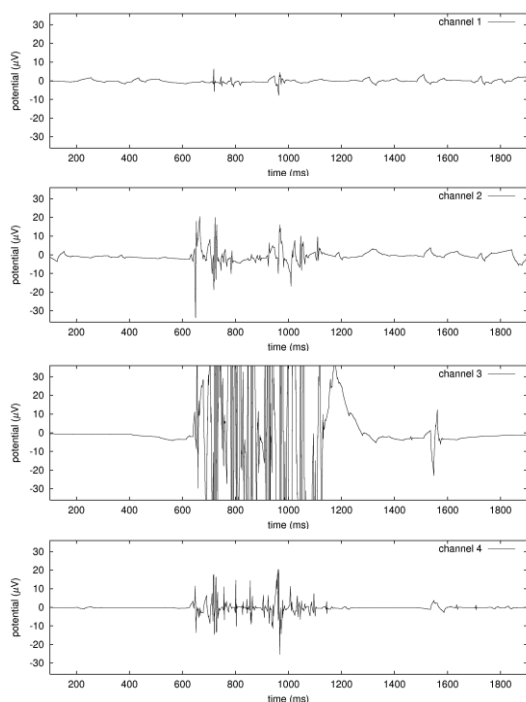


図 3. 図 2 の波形を信号強度方向に拡大したもの

しかしながら、図 3 のように信号を拡大してみると、ほぼ信号なしと捉えられていたチャンネル 1 や 2 に、2 回の[イ]の発声に呼応した特徴が現れていることがわかる。とはいえ、特にチャンネル 1 に現れた特徴は、ノイズとの区別がつかないほどの微弱な信号強度しか持たないものである。周波数変化は大きいよ

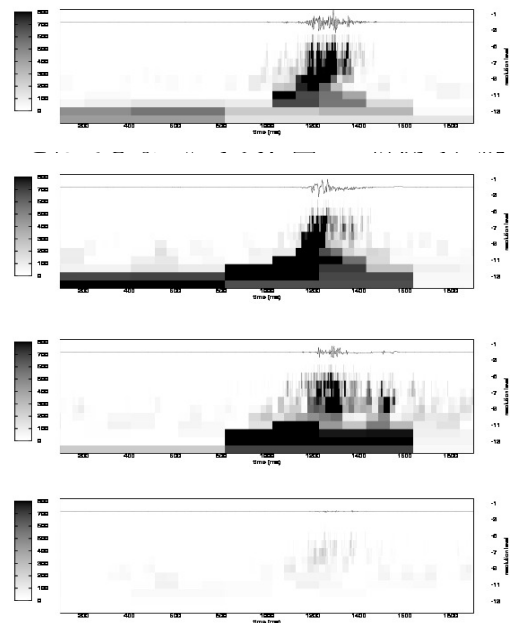


図 4. 母音[ウ]の表面筋電信号のウェーブレット解析結果

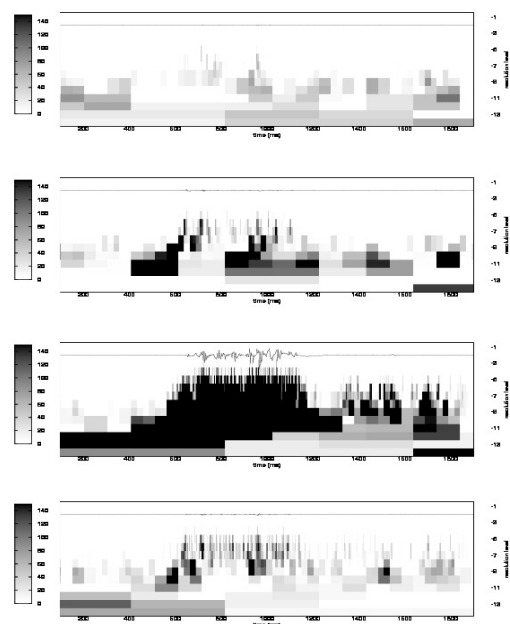


図 5. 図 2 の信号のウェーブレット解析結果

このように短時間での周波数変動を伴う信号の解析にはウェーブレット解析が適している。図 4 に母音[ウ]と発声した際の表面筋電信号をウェーブレット解析した結果を示す。横軸が時間、縦軸が周波数帯域（レベル）、濃淡がウェーブレット係数の強弱を表す。参考として、図の上部に筋電信号波形を掲載している。この例のように、筋電波形に強い信号が観測されるよりも前から信号が強い領域が低周波数領域に発生し、高周波数領域へと移行していく様子が観測される。逆に終了時には逆に高域から低域への移行が

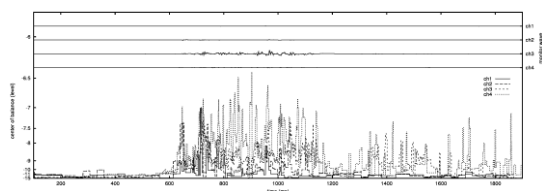


図 6. 連続母音[イ]の重心推移

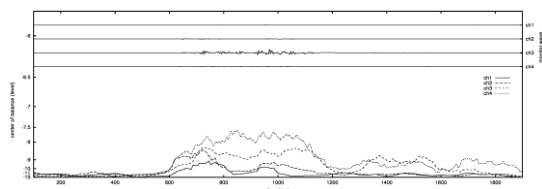


図 7. 64ms 幅で平均化した[イ]の重心推移

生じる。この傾向は 2 音間についても存在し、先行音の終了と後続音の開始との特徴が重なったような変化が見られる。

こうした特徴を扱う際、ウェーブレット係数値を直接に用いたのでは信号強度で扱う場合の微弱信号と同じ問題が生じる。そこで信号の強い領域の推移の特徴を捉えるための手法として、ウェーブレット係数の重心推移法 (Center-of-Balance Transition; CoBT) と呼ぶ手法を提案した。重心推移法とは、特定時間におけるウェーブレット係数を切り出し、レベル差を距離、ウェーブレット係数値を重さとして周波数軸上の重心位置を求め、その時間推移によって筋の活動状況を捉える手法である。図 2 の連続母音[イ]の重心推移を図 6 に示す。

重心推移の時間軸方向の解像度は、離散ウェーブレット解析の性質上、サンプリングレートの 1/2 である。図 6 の例からもわかるとおり、この解像度は大局的な変化を捉えようとする場合には少々詳細すぎる傾向がある。そこで、重心推移を 64ms 幅の移動平均で平滑化した結果が図 7 である。重心推移では、ノイズと区別がつかないほどに微弱であったチャンネル 1 の信号からも、信号が強かったチャンネル 3 と十分に比較可能なレベルで、2 回の[イ]の発声に伴う山として特徴を観測できていることがわかる。

自然な連続発声においては、変化が緩やかである上に、アクセントのような発声の強弱に呼応して信号の強弱も変動する。アクセント情報が得られることは良いことであるが、これにより発声に変化した位置がどこであるのかを同定するのが難しく、連続黙声認識の障壁の一つとなっている。重心推移は、信号強度をそのまま用いるのに比べれば、アクセントなどの強度の違いによる影響を受けにくい性質を持つ。そこで本研究課題では、この重心推移を活用して、黙声連続 2 母音の発声変化位置 (先行音の発声開始位置、後続

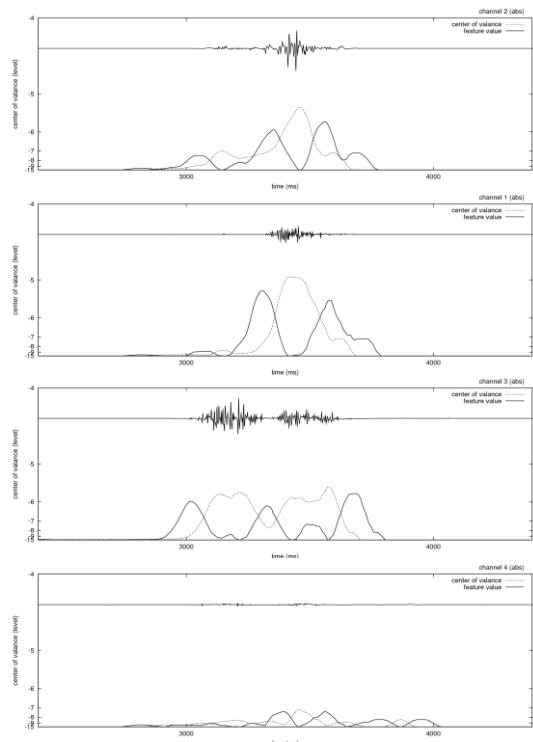


図 8. 連続母音[イオ]の重心推移から求めた変化位置候補の特徴量

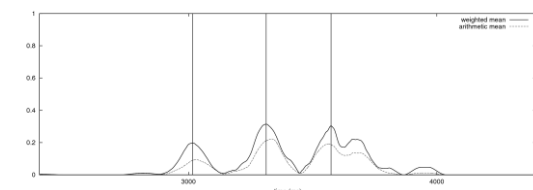


図 9. 連続母音[イオ]の発声変化位置

音への変化位置、後続音の発声終了位置) を同定することを試みた。

先に述べたように、ウェーブレット解析結果において、2 音の切り替わり位置では先行音の終了部の特徴と後続音の開始部の特徴とが重なったような変化を示す。この性質は、平滑化した重心推移においては谷を形成する傾向が強いということを意味する。そこで、この谷を同定することで、発声変化位置の候補を抽出することとした。64ms 幅で平滑化した結果でもまだ細かい変動が多く存在するので、64ms 幅で平滑化した後にさらに 32ms 幅で平滑化するという 2 段階の平滑化を施したものをを用いた。これに対し、ある種の包絡線を求め、平滑化後の重心値と包絡線との差を出力する。この結果は、値が大きいほど重心推移の谷が深い、すなわち、変化点らしさが大きいと見ることができる。連続母音[イオ]におけるこの特徴量の例を図 8 に示す。本研究課題では、各筋で得られた値の重み付き平均を取り、その値が設定した閾値を超える連続区間の中で最大値を示した位置

を変化点候補として抽出した。変化点候補は、発声終了後の残存活動に対して過剰に抽出される傾向があったため、変化点候補によって切り出された区間の筋活動量（RMS 値を利用した）が閾値未満であった場合には無効な候補として絞り込みを行った。連続母音[イオ]の例において抽出された発声変化点候補を図9に示す。本手法により、2連続母音の発声変化点の平均の再現率は92.5%、適合率は88.8%、F 値は90.1 を得た。

検出した発声変化点の情報に基づいて、その前後の筋活動を活用して2母音の認識を行うことも試みたが、従来の単母音認識と比べてあまり良い結果は得られなかった。検討を行ったところ、自然な発声においては、ある母音の発声中であっても、次に発声する音を意識した口唇の動きがなされていることがわかった。重心推移において先行音の発声が終了する前に後続音の影響が出始めているのも、このことが原因であると考えられる。これはやっかいな問題ではあるが、発声から認識までのタイムラグを縮小できる可能性も含むため、今後の研究の重要な課題の一つとする予定である。

黙声子音の認識についての研究では、各筋の活動量だけではなく、活動増減に関する協調動作のタイミングの重要性が高いことがわかってきた。そこで、ウェーブレット係数の重心値の差分推移から筋活動の増減のトレンドを捉える手法を提案し、口唇の動きに特徴が出やすい両唇音を対象に分析を進めている。子音認識のための計測位置については2か所を試験的に追加してみたが、内1か所については効果が見込めなかった。もう1か所については、有効性の評価を進めている。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計2件）

- ① 宮武一志、永井秀利、中村貞吾、筋活動量を考慮したウェーブレット係数の重心推移法に基づく連続黙声発話の変化点検出、電子情報通信学会技術研究報告、査読無、Vol.112、No222、2012、pp.69-74
- ② 永井秀利、宮武一志、中村貞吾、黙声認識のための口唇周辺表面筋電波形におけるウェーブレット係数の重心推移法に基づく特徴、FIT2012 講演論文集、査読有、2011、pp.41-46

〔学会発表〕（計8件）

- ① 永井秀利、金井田早紀、宮武一志、中村貞吾、黙声認識における重心推移法に基づく差分推移を活用した発声時筋活動変化の抽出手法、HCG シンポジウム2012、

II-6-4、2012年12月11日、くまもと森都心プラザ

- ② 永井秀利、金井田早紀、宮武一志、中村貞吾、黙声認識のためのウェーブレット係数の重心値の差分推移に基づく表面筋電信号の分析、第65回電気関係学会九州支部連合大会、06-1A-16、2012年9月24日、長崎大学
- ③ 宮武一志、永井秀利、中村貞吾、表面筋電波形におけるウェーブレット係数の重心推移に基づく連続黙声発話の変化点検出、火の国情報シンポジウム2012、C-6-3、2012年3月16日、九州工業大学
- ④ 宮武一志、永井秀利、中村貞吾、表面筋電波形のウェーブレット係数に基づく2黙声間発声変化点検出の精度向上手法、第64回電気関係学会九州支部連合大会、05-2A-01、2011年9月27日、佐賀大学
- ⑤ 宮武一志、永井秀利、中村貞吾、表面筋電波形のウェーブレット係数に基づく黙声の発声変化点検出、FIT2011、J-043、2011年9月7日、函館短期大学
- ⑥ 永井秀利、宮武一志、岩崎亘、中村貞吾、黙声認識のためのウェーブレット係数の重心推移に基づく表面筋電波形の特徴分析手法、火の国情報シンポジウム2011、B-6-3、2011年3月9日、福岡大学
- ⑦ 永井秀利、宮武一志、岩崎亘、中村貞吾、ウェーブレット解析を用いた黙声認識のための表面筋電波形の特徴分析ー2単音間の推移の特徴を中心として、第63回電気関係学会九州支部連合大会、03-2P-07、2010年9月26日、九州産業大学
- ⑧ 宮武一志、岩崎亘、永井秀利、中村貞吾、ウェーブレット解析を用いた黙声認識のための表面筋電波形の特徴分析ー発声開始・終了点の特徴を中心として、第63回電気関係学会九州支部連合大会、03-2P-06、2010年9月26日、九州産業大学

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