

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

宮 武 一 志^{†1} 永 井 秀 利^{†1} 中 村 貞 吾^{†1}

我々は、筋電信号を用いた黙声（いわゆる口パク）認識の研究を行っている。従来の研究では、表面筋電信号のウェーブレット係数の重心値の推移によって特徴を捉える手法（重心推移法）を提案し、その応用として、連続黙声の発声変化点の検出を試みた。従来の発声変化点検出では、いずれの音の発声でも強く活動する傾向を持つ筋において発声変化点で生じる重心値の低下のみを活用していたが、十分に満足できる検出精度は得られなかった。そこで本研究では、他の筋における重心値推移の特徴も合わせて変化点らしさを評価することにより、検出精度の向上を目指した。本稿ではその結果について述べる。

Detection of Phoneme Boundary in Inaudible Speech Utterance based on Center-of-Balance Transition of Wavelet Factors on Surface EMG

KAZUSHI MIYATAKE,^{†1} HIDETOSHI NAGAI^{†1}
and TEIGO NAKAMURA^{†1}

We are working on inaudible speech recognition based on surface EMG. In our previous work, we proposed a method for extracting features of sEMG by "Center-of-Balance Transition" (CoBT) of wavelet factors. With this method, we tried to detect phoneme boundary in inaudible utterance. The detection method on our earlier research uses a sEMG of a muscle working strongly on any kind of vowels. On such sEMG, Center-of-Balance value is decreasing at boundary. We utilized the feature for the method, but we couldn't get enough precision. In this paper, we improve the detection method with suitability for boundary, which is estimated by features on a set of CoBTs.

1. はじめに

我々は声を出さずに発声された内容を口唇周辺から頸部の表面筋電信号に基づいて認識する研究を行っている。我々はこれを筋電による黙声認識と呼ぶ。本技術によって、普通に話す場合と同様に口を動かすことにより可聴音を発することなく計算機への自然言語入力が可能となる。これにより、機密性の高い音声入力やマイク入力困難時の音声認識支援に役立つほか、新たな音声入力インターフェース、喉頭切除で声を失った人の発声代行など、様々な活用が考えられる。

従来の研究^{1)–4)}では、可聴音の発声は一切ない場合であっても可聴音を発声した際と類似した表面筋電波形が観測されることを確認の上、黙声単母音における発声開始情報を用いた認識を試みて良好な結果を得ている。そこで次に孤立単音認識から連続音認識への発展が考えられるが、連続音では一音ごとにアクセントや抑揚などによる信号強度の変動が存在し、そのままでは特徴が安定せず、十分な認識精度を得ることは困難である。そこで本研究では、発声区間を切り分けた上で正規化などを施し、個々の区間の特徴からその区間の発声区間内容を認識することで連続音認識を行うこととする。しかしながら、従来の信号強度に基づく特徴量のみでは十分な精度での区間切り分けは難しい。これは自然な連続発声では口唇形状の変化が音から音へと推移しつづなされるため、直前の発声での活動が後続の発声の頭にまで影響を及ぼすことが原因である。また、アクセントに伴って生じる筋活動の強弱の変動なども変化点検出を難しくする。1音ずつはっきりと区別しつつ発声することにより変化点を検出することは可能かもしれないが、そのような発声方法は負担が大きく、発声を長時間続けることは非常に難しい。したがって信号強度に依らない特徴であり、発声の強弱に対して比較的鈍感な特徴を抽出する必要がある。

そこで従来の研究では、信号強度に依らない特徴として表面筋電波形のウェーブレット係数の重心推移における特徴を利用し、黙声による連続2母音発声時の発声変化点検出を試みた。その際一部の筋電信号のみを用いて検出を行なっているが、これはいずれの発声でも強く活動する傾向がある筋が、発声変化点において一時的な重心値の低下を見せる点を活用したためである。しかしこの特徴は筋電信号の微小な変化に大きく影響を受けるため、安定した発声変化点の検出は難しかった。

^{†1} 九州工業大学
Kyushu Institute of Technology

2 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

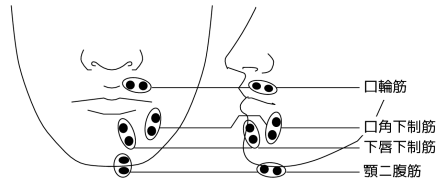


図1 電極装着位置の概略図

そこで本稿では、他の筋においては先行音か後続音かのいずれかのみで活動が強く、変化点において重心値の単純な上昇や下降のみが生じる傾向を併せて活用することとし、発声変化点検出を行った。

2. 筋電信号について

本研究で目的としているのは発声内容の認識であって、口唇の形状や顎の開放角度などを正確に捉えることではない。したがって発声に関わるすべての筋を測定しなくても、識別したい発声の特徴を判断できるような相対関係が得られれば十分である。そこで本研究では利用上の負荷の小ささを鑑み、表面筋電に基づいて認識を行うものとしている。

本稿における計測対象は、従来の研究¹⁾ で用いた口輪筋、口角下制筋、下唇下制筋、顎二腹筋の4筋とし、4チャンネルのアクティブ電極を使い計測した。図1に電極装着位置と各位置の目標筋とを示す。電極装着位置の皮膚をアルコールで清拭した後、目標としている筋に対してそれぞれ計測器のチャンネルを割り当て、皮膚から浮かないように十分注意して装着した。

計測時の発声に関しては、静かに口を閉じた状態から始め、連続する二つの母音を発声した後、急激過ぎない程度に元の状態に復帰するという方法で行った。このとき、口唇形状を無理に強調することは避け、自然な発声となるよう心掛けた。筋電信号データは発声過程ごとに5秒間とし、解像度12bit、周期 $200\mu\text{s}$ (5,000Hz) でサンプリングした。得られた計測波形に対し、ウェーブレット縮退を利用したノイズ低減手法⁵⁾ を適用したものを最終的に波形データとして収集した。

3. ウェーブレット解析結果に基づく特徴分析

獲得した波形データに対し、Daubechies'N=2を用いてウェーブレット解析し、特徴分析を行った。

[イオ]を発声した場合の表面筋電波形のウェーブレット解析結果の例を図2に示す。channel1から順に、口輪筋、口角下制筋、下唇下制筋、顎二腹筋を目標とした計測位置での結果である。図の横軸は時間、縦軸は多重解像度解析における周波数帯域のレベルを表し、濃淡はウェーブレット係数の大きさを表している。なお図上部の波形は解析対象の筋電信号波形を参考として示したものである。

発声開始部ではウェーブレット解析結果における低い周波数帯域から徐々に中、高域へと信号が強い領域が推移していく傾向が見られる。信号が強い領域は実際の発声よりも早い時点に発生しているため、これは発声に向けての準備動作が反映されたものであると考えられる。また発声終了部は逆に活動が強い領域が高域から低域へと急激に推移する傾向が見られる。そして発声変化点とみられる位置では、このような特徴が前後2音での特徴の重なりを伴いつつ現れる傾向が多く見られた。つまり発声変化点では前の音の終了に伴う強領域の下降傾向と次の音の強領域の上昇傾向が重なったような推移が見られる。

同様の特徴が同一母音の連続発声に関しても見られた。長音を発声した場合と比較すると、同一母音の方は長音の方には見られない上記特徴が見られる。しかし人は同一母音の連続発声と長音発声をさほど明確に区別してはいないため、このような両発声での差異が必ず発生するわけではない。

これらの特徴は強弱の差はあるがどの筋に関しても同じように観測されることが多く、特に下唇下制筋において最も顕著に現れた。これは発声開始の際に口唇を開く必要があり下唇下制筋が大きく動くためであると考えられる。他の筋に関しては先行音か後続音かのいずれかのみで活動が強くなる場合もある。これはその筋を使用しないで発声を行う、または先行音で用いた筋を脱力することによって発声を行う場合があるからである。

こうした周波数帯域の強領域の推移に関する特徴は、計測対象外の筋電信号の混入が影響している可能性を完全には否定しないが、主たる活動と見なせる部分の信号との相関が十分存在していることから、特徴として活用することに支障はないと考えている。

4. 特徴のパラメータ化

第3章で述べた時間-周波数平面上の特徴を変化点検出に活用できるように定式化する必要がある。そこで我々は信号が強い領域の推移や散らばりに着目し、ウェーブレット係数の重心という概念を定義する。これは特定時点における各レベルのウェーブレット係数を重みとし、周波数(レベル)軸上の重心の位置を求めたものである。このとき、処理対象とする最も低い周波数帯域のレベルを $-L$ とし、時刻 t を含む領域のウェーブレット係数を

3 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

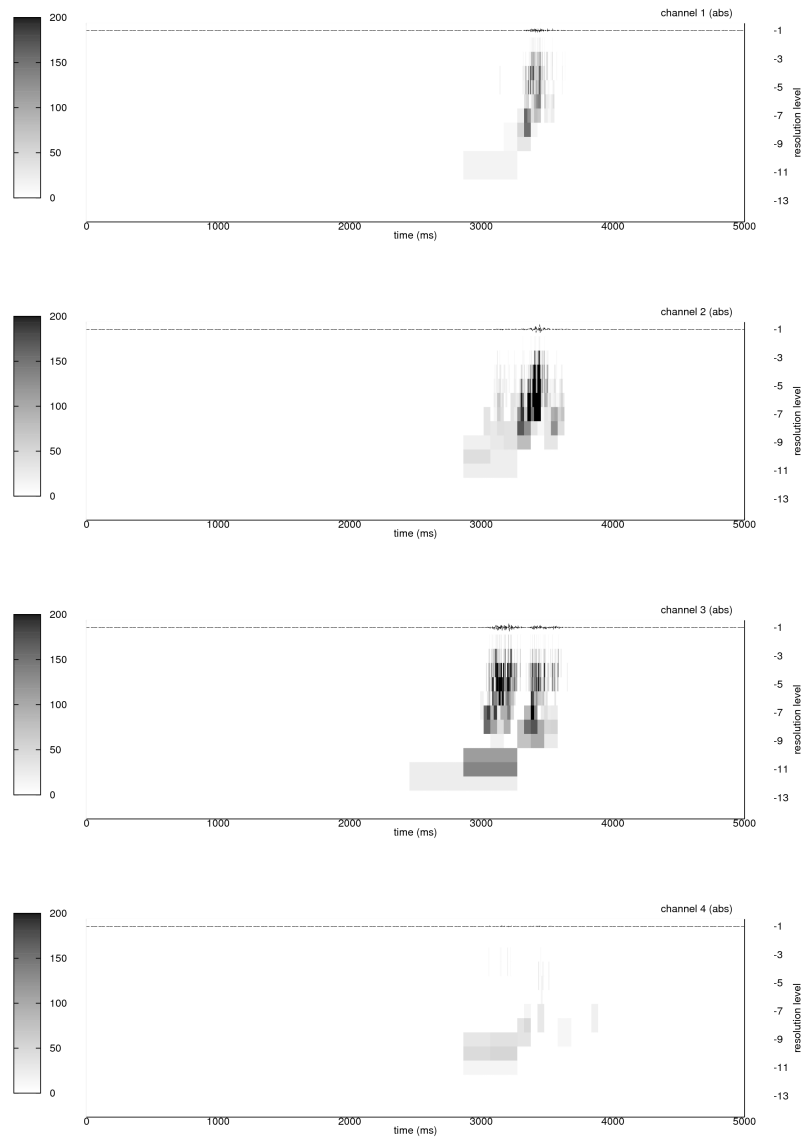


図 2 連続母音 [イオ] のウェーブレット解析結果

$w_{-1}(t), \dots, w_{-L}(t)$ とすると、重心値 $CoB(t)$ は次式となる。

$$CoB(t) = \begin{cases} 0 & (if \sum_{k=1}^L |w_{-k}(t)| = 0) \\ \frac{1}{\sum_{k=1}^L |w_{-k}(t)|} \sum_{k=1}^L |w_{-k}(t)| \cdot (2^{L-k} - 1) & (otherwise) \end{cases}$$

この重心値が時間軸によってどのように推移するかによって筋電信号の特徴を捉える手法を、我々は重心推移法⁶⁾ (Center-of-Balance Transition) と呼ぶ。図 2 のウェーブレット係数の重心推移を図 3 に示す。

重心値は筋活動が行われている部分では高い値が出現するが、表面筋電信号の強さとは必ずしも呼応しない。つまり信号強度に依らない特徴を持つため、発声の強弱に対して比較的鈍感な特徴であると言える。

この重心推移法により、筋活動の細かい変化を追うことができるが、図 3 に示したような値のままでは変化が細かく、第 3 章で述べたような大局的な変化を追いつらい。よって本研究では発声に伴う緩やかな変化を捉えるため、平滑化を行う。平滑化幅は、計測した種々の発声を調査した結果として、50~75ms 程度の幅で重心推移の平滑化を行うのが妥当だと判断した。これより短いと変動が多すぎて主要な変化点を捉えづらくなり、逆に長いと信号を均しすぎて傾向が失われてしまう傾向がある。図 4 に図 3 で示した重心推移を時間幅 62ms で平滑化した結果を示す。

5. 発声変化点検出手法

発声変化点には発声開始部、発声終了部、2 音間の発声変化部分の 3 種類が考えられる。第 3 章で述べた特徴から考えると、発声開始部では活動の強い領域が低域から高域へと推移し、発声終了部では逆に高域から低域へと推移する。重心推移でも同様の変化が見られると考えられ、それぞれ重心推移の上昇または下降という傾向が生ずる。また 2 音間発声変化点では、高域におけるウェーブレット係数が減少から増加という変化が生じるため、重心推移にも同様の変化が見られるはずである。つまり発声変化点において一時的な重心値の減少が見られる。

そこで本研究では、重心推移波形における上昇、下降傾向と一時的な重心値の減少傾向を特徴量化することによって発声変化点を捉えることとする。方法としては重心推移波形の極大部分を繋ぐような包絡線を作成し、その線と重心推移波形との差分を取る。すると重心値が増大、減少、一時的な減少が生じた部分において差分値が大きくなることが予想される。差分値を求める具体的な手順としては、まず特定の幅の窓をずらしながら、その窓区間内の

4 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

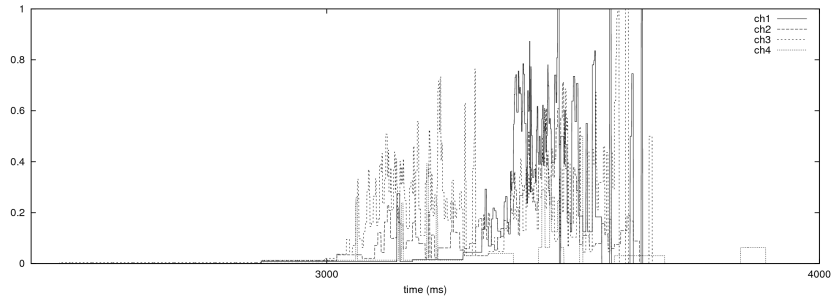


図 3 連続発声 [イオ] の重心推移

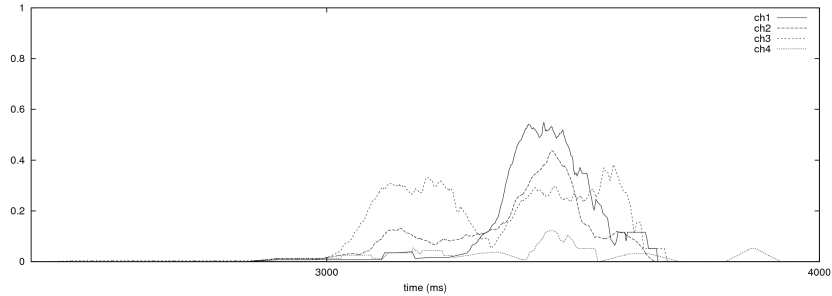


図 4 連続発声 [イオ] の重心推移 (平滑化後)

最大値を求める。その求めた最大値を特定窓幅内の中心点での値とし、全ての時間軸でそれぞれ求める。このとき窓幅は極小部分を超える時間幅である必要があるので 1 音の時間幅から考えて 100ms とした。次に得られた値と重心推移との差分をとることによって、その筋における重心推移波形の上昇、下降傾向と一時的な減少傾向を特徴量化した値となる。各筋で求めたこの値を平均することによって発声変化点らしさを表す特徴量が得られる。よってこの値が一定の閾値を超えた連続区間内での最大値を示す位置を発声変化点として検出する。図 6 における縦線は、本手法を適用し検出した発声変化点である。図 5 に図 4 で示した重心推移の差分値を求めた結果を示す。実線がその筋における重心推移であり、破線が差分値である。その差分値を各筋で求め結果を平均したもの図 6 に示し、これを変化点検出に使用する。このそれぞれの差分値を合計した値は発声変化点である可能性が高い部分においてより大きくなることが考えられるため、この値が大きくなる部分を捉えることによって発声変化点を検出する。

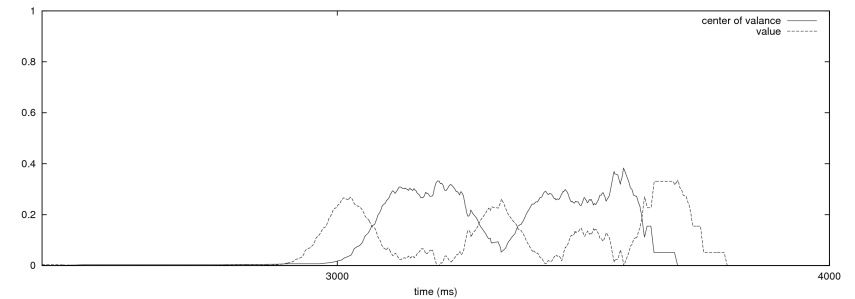
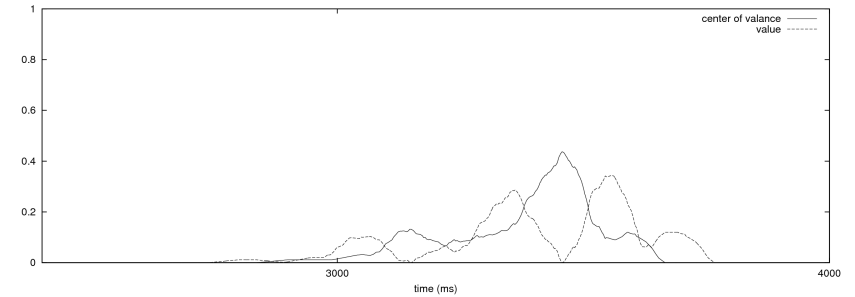
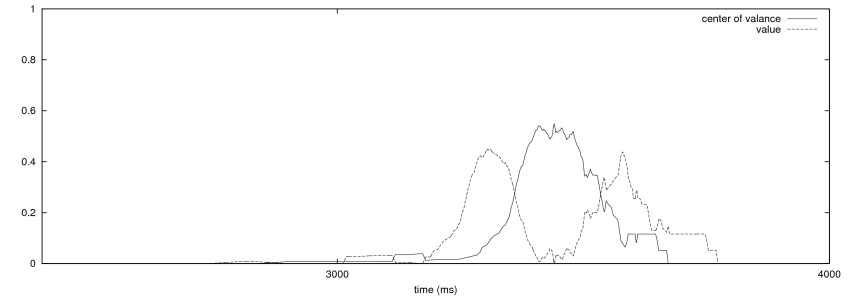


図 5 連続発声 [イオ] の重心推移から求めた特徴量

5 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

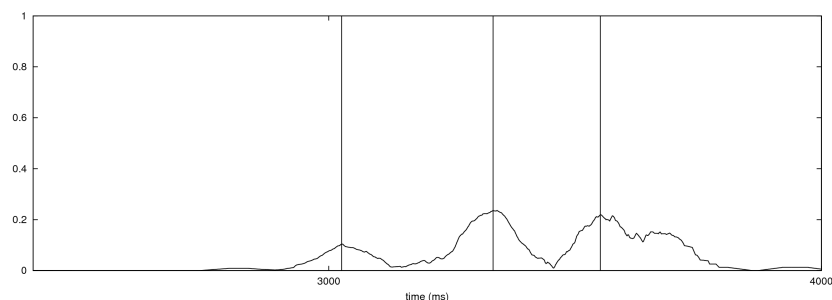


図 6 連続発声 [イオ] の重心推移から求めた特徴量の平均

6. 実験結果

前述した方法を用いて発声変化点の検出を試みた。実験には、各母音の組み合わせで 2 音の連続音 25 種類を各 8 回測定したものをを用いた。自然な連続発声では口唇形状を変化させながら発声を行うため、発声変化点というものは一点に定まるものではない。そのため今回の実験で発声変化点の検出を行っているかどうかは、発声内容、及び 4 チャンネルの筋電波形とウェーブレット解析結果から人手で変化点を推定して正誤を定めた。再現率と適合率の結果を表 1、表 2 に示す。

本手法による平均の再現率は 84.0%、適合率は 81.7%、F 値は 82.9 となった。従来の研究⁷⁾ における手法では、平均の再現率は 82.5%、適合率は 66.6%、F 値は 73.7 であった。従来の手法では 2 音間発声変化点のみを捉えていたため、単純に結果を比較することはできない。しかし今回の手法はデータ内に含まれる全ての変化点を検出したにもかかわらず、従来の手法よりも再現率、適合率ともに向上する結果となった。

従来の研究⁸⁾ では、[オ] を先行音とした場合における検出率の低下が再現率を下げる要因として挙げられていた。これは、下唇下制筋において、[オ] という発音は他の母音よりもウェーブレット係数の値が比較的小さくなる傾向があるためである。先行音のウェーブレット係数が小さい場合、後続音の準備活動により発生する低周波領域のウェーブレット係数が、重心値の上昇を抑制する。そのため 2 音間の発声変化点において期待した特徴が重心推移波形に現れず、変化点を捉えることが難しくなる。本実験では下唇下制筋だけではなく、他の筋の特徴も用いて発声変化点検出を行ったため、2 音間の発声変化点において特徴を強調することが可能となり、高い再現率を得ることができた。

表 1 変化点検出結果（再現率 [%]）

		後続音					
		あ	い	う	え	お	平均
先行音	あ	54.2	83.3	70.8	75.0	79.2	72.5
	い	58.3	91.7	95.8	83.3	95.8	85.0
	う	70.8	91.7	100	87.5	95.8	89.2
	え	62.5	95.8	91.7	79.2	95.8	85.0
	お	79.2	100	91.7	83.3	87.5	88.3
	平均	65.0	92.5	90.0	81.7	90.8	84.0

表 2 変化点検出結果（適合率 [%]）

		後続音					
		あ	い	う	え	お	平均
先行音	あ	92.9	83.3	85.0	81.8	79.2	84.4
	い	100	84.6	88.5	87.0	82.1	88.4
	う	85.0	73.3	75.0	80.8	74.2	77.7
	え	88.2	76.7	81.5	79.2	76.7	80.4
	お	79.2	77.4	78.6	83.3	70.0	77.7
	平均	89.1	79.1	81.7	82.4	76.4	81.7

従来に比べて特に適合率が高くなったが、これは、従来は誤検出としていた発声終了部の発声変化点を捉えるようにしたことも要因の一つとなっている。従来手法では、まず最初に重心値が一定の閾値を超えた連続区間を発声区間として同定しておいた上で、その区間内に 2 音間の発声変化点が必ず存在するものとして変化点を捉えるようにしていた。その際、発声終了後の筋活動が大きかったために、その部分までを発声区間として捉えてしまい、実際の発声区間と発声終了後の筋活動との間を発声変化点として捉える場合があった。本手法と比較するなら、その位置を発声変化点としても間違いとは言えないが、従来手法の対象は 2 音間であり、発声終了点については対象としていない変化点が偶然検出できたに過ぎないため、誤った検出として扱っていた。

また発声によっては、2 音間での重心値低下が非常に大きくて 1 音目のみを発声区間として同定（2 音目は発声終了後の筋活動と認識）してしまい、1 音目の途中に無理矢理に誤った変化点を検出する場合もあった。よって本研究では、予め同定した発声区間に依存する手法を用いないようにすることで、発声区間同定の誤りに起因した誤検出を減らすことができた。

再現率を低下させた要因としては、開始部、終了部での発声変化点が検出できていない場

6 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

合が多々あったためである。今回用いた4筋の内で先行音または後続音しか活動しない筋が含まれる発声するとき、他の筋で特徴量が得られても、その活動しない筋の特徴量によって全体が均されてしまい、検出に失敗したと考えられる。これは[A]を先行音、後続音とした場合の再現率が低いことから言える。[A]を発声する際には主に下唇下制筋を用いるため他の筋の活動が少なく、期待した特徴が得られなかったため検出が難しくなった。特に[A]が後続音となった場合は先行音で開いていた口のまま発声することが可能なため、新たな筋活動を必要とせず、脱力動作のみでの発声が可能となる。そのため期待した特徴が出現しづらく、他の発声の場合に比べて特に再現率が低い結果になった。逆に先行音でのみ強く活動する筋と後続音でのみ強く活動する筋と両方で強く活動する筋がバランスよく含まれている発声の場合は、全ての場合で発声変化点検出を高い精度で行うことができた。

適合率を低下させた要因として、発声終了後の発声以外の筋活動部分を変化点として過剰に検出してしまった場合がある。これは発声終了後に口唇形状を戻そうとするときなどの筋活動により発生する。発声変化点としては誤検出であるが、筋活動という点で考えると変化点であるため、本手法では検出してしまう。この区間は我々が必要とする発声区間ではないため、後処理によって排除する必要がある。他には変化部分でのゆらぎによって短い時間幅で多数検出してしまった場合がある。これも現在の手法のみで改善するのは困難であるので、発声のピッチ情報などを用いて検出結果から正しい発声変化点を選択するといった処理が必要である。

7. おわりに

ウェーブレット解析結果の重心推移を特徴量として、2音間の母音間の発声変化点検出を行った。再現率、適合率ともに、従来の発声変化点検出結果に比べて良好であった。これは下唇下制筋以外の筋に現れる特徴を利用したため、従来特徴が弱まっていた部分での特徴が強調されたためである。特に従来は検出が難しかった一部母音において高い検出結果を得ることができた。しかし活動しない他の筋の影響によって特徴量が均されてしまい、検出が困難になる場合も存在した。そのため特徴量化にはもう一工夫必要であり、本手法で単純に4筋全ての特徴量を平均したものをを用いていた部分を、活動する筋のみを使用するなどの改善が考えられる。

今後は更に検出精度を高めていくとともに、子音を含む場合や一音ごとに切り出した部分に対する認識実験を進めていきたい。

謝辞 本研究は科研費（課題番号：22500157）の助成を受けたものである。謹んで感謝の意を表する。

参 考 文 献

- 1) 永井秀利, 中山浩之, 中村貞吾, 野村浩郷: "筋電に基づく黙声認識におけるニューラルネットワークを用いた母音認識", 第57回電気関係学会九州支部大会 12-1P-05 (2004).
- 2) 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", 電子情報通信学会技術報告 Vol.102, No.688, pp.25-32 (2003).
- 3) 永井秀利, 中村貞吾, 野村浩郷: "無発声ないし微発声音声認識への活用を目的とした表面筋電波形の調査", 情報処理学会第65回全国大会講演論文集 2F-7 (2003).
- 4) 永井秀利, 宇土由紀, 中村貞吾: "口裂周辺の筋電信号を用いた黙声日本語単母音認識のための認識パラメータ獲得手法", 第8回情報科学技術フォーラム (FIT2009) 論文集 E-053 (2009).
- 5) 永井秀利, 中村貞吾, 野村浩郷: "無発声ないし微発声音声認識のための表面筋電波形からのノイズ低減手法", 情報処理学会九州支部大会「火の国情報シンポジウム 2003」, pp.1-8 (2003).
- 6) 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に基づく特徴", 第10回情報科学技術フォーラム「FIT2011」 R.J-006 (2011).
- 7) 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", 第10回情報科学技術フォーラム「FIT2011」 J-043 (2011).
- 8) 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).