

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

Detection of Phoneme Boundary in Inaudible Utterance
by Wavelet Factors of Surface EMG宮武 一志[†]
Kazushi Miyatake永井 秀利[†]
Hidetoshi Nagai中村 貞吾[†]
Teigo Nakamura

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。本技術では、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。

従来研究では、黙声単母音における発声開始情報を用いた認識を試み、良好な結果を得ている。孤立単音認識から連続音認識へと発展させる場合、発声の変化点を捉え、発声区間を切り分ける必要がある。しかし、従来研究で用いてきた信号強度に基づく特徴量のみでは発声変化点検出は難しい。これは、自然な連続発声の場合、音から音へと推移しつつ口唇形状の変化がなされるため、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ためである。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。

そこで本研究では、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、黙声による連続発声時の発声変化点検出を行った。本稿ではその結果について述べる。

2. 筋電計測について

計測対象は、従来研究 [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、電極装着位置の皮膚をアルコールで清拭した後、4チャンネルのアクティブ電極を使い計測した。

表面筋電は解像度 12bit、周期 $200\mu\text{s}$ ($5,000\text{Hz}$) でサンプリングした。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。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 [2] 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] を発声した場合の下唇下制筋の表面筋電波形のウェーブレット解析結果の例を図1に示す。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。

ウェーブレット解析結果の発声開始部では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で発生しているため、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言える。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。

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、今回の計測対象である連続2母音においては下唇下制筋に最も顕著に現れた。これは、発声開始の際に口唇を開く必要があることで、下唇下制筋が大きく動くためと考えられる。よって、本稿の発声変化点検出は下唇下制筋の筋電信号を用いて行うこととした。

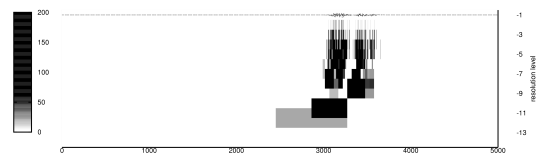


図1: 連続発声 [イオ] の下唇下制筋のウェーブレット解析結果

4. 特徴のパラメータ化

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。そのために本研究では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 $w_{-1}(t), \dots, w_{-L}(t)$ とすると、重心値 $CoB(t)$ は次式となる。

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

図1のウェーブレット係数の重心値が時間軸に沿って推移していく様子を図2に示す。このように、重心値の推移から特徴を捉える手法を我々は重心推移法 (Center-of-Balance Transition) と呼ぶ。この手法により、第3

[†]九州工業大学, Kyushu Institute of Technology

章で述べたような変化点の特徴を追うことができるが、図2に示したような値のままでは変化が細かく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、62msの時間幅で平滑化を行った。その結果を図3に示す。

5. 発声変化点の抽出手法

3章で述べた特徴から考えると、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、重心の推移にも同様の変化が見られるはずである。そこで本研究では、重心推移における発声区間の極小部分を検出することにより発声変化点を捉える。方法としては、発声区間の極大部分を繋ぐような包絡線を作成し、その線と重心推移との差分を取る。すると極小部分では差分値が大きくなるため、その部分を捉えることによって極小部分を検出する。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、その窓区間の最大値をそれぞれ求める。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。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り、その値が閾値を超えた部分での最大値を示す位置を変化点として検出する。図3における縦線は、この手法を使用して検出した発声変化点である。

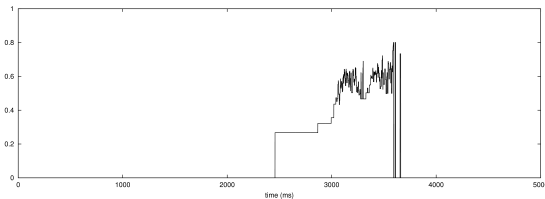


図2: 連続発声 [イオ] の下唇下制筋の重心推移

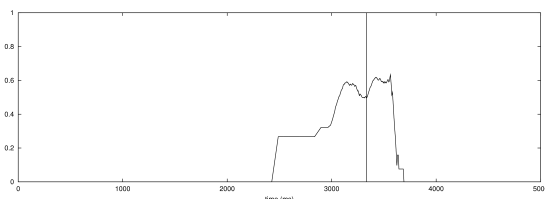


図3: 連続発声 [イオ] の下唇下制筋の重心推移 (平滑化後)

6. 実験結果

下唇下制筋の筋電信号に対し、前述した方法を用いて発声変化点の検出を試みた。実験対象は、各母音の組み合わせで2音の連続音25種類を各8回測定したものである。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、人手で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変化点を推定し、正誤を定めた。再現率と適合率の結果を表1と表2に示す。

結果として、平均の再現率は82.0%、適合率は66.1%となり、再現率に比べ適合率が低いという結果を得た。適合率低下の原因としては、発声後に口唇形状を戻そうとするときなどの発声以外の筋活動を3音目として捉え変化点を検出したことが挙げられる。これは本実

験で期待した検出ではないが、発声変化点検出という意味では正しい検出結果である。

また、再現率を下げる要因として、先行音が[オ]の場合における検出結果が悪いことが挙げられる。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ことが原因である。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、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。また、1音目に[オ]を発声した場合、2音目には口唇形状を開いた状態から閉じる状態へと移りながら発声するため2音目における筋活動が下唇下制筋に現れず、特徴が観測され難い。これらの要因が、[オ]が先行音の場合における検出率の低下に繋がったと考える。

表1: 変化点検出結果 (再現率 [%])

		後続音					平均
		あ	い	う	え	お	
先行音	あ	87.5	75.0	75.0	87.5	100	85.0
	い	62.5	87.5	100	100	100	90.0
	う	75.0	87.5	87.5	87.5	62.5	80.0
	え	62.5	87.5	100	100	100	90.0
	お	62.5	75.0	62.5	62.5	62.5	65.0
平均		70.0	82.5	85.0	87.5	85.0	82.0

表2: 変化点検出結果 (適合率 [%])

		後続音					平均
		あ	い	う	え	お	
先行音	あ	57.1	66.7	75.0	70.0	66.7	67.1
	い	50.0	61.5	57.1	61.5	80.0	62.0
	う	66.7	63.7	76.9	77.8	62.5	69.5
	え	62.5	72.7	57.1	100	66.7	71.8
	お	50.0	75.0	60.0	55.6	60.0	60.1
平均		57.3	67.9	65.2	73.0	67.2	66.1

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音の母音間の発声変化点検出を行った。再現率に関しては良好だが、適合率が低いという結果を得た。これは本実験で対象とした2音間の変化以外の発声後の変化まで多く検出してしまったことが原因であり、前述したように発声変化点検出としては良好な結果を得ていると言える。つまり、これは本手法が発声変化点を捉える一手法として有効なものであることを示している。

今後は、更に検出精度を高めていくとともに、3音以上の連続音への適用を行っていきたい。そのために、発声変化点特徴が顕著に現れるように重心値の上昇を抑制してしまう低周波成分の除去や、他筋肉との重心推移の相対関係などを利用していきたいと考えている。

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。

参考文献

- [1] 永井 他: “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実験”, 第57回電気関係学会九州支部大会 12-1P-05(2004)
- [2] 永井 他: “無発声ないし微弱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”, 情報処理学会第65回全国大会講演論文集, 2F-7(2007)