E-053

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

Methods to Acquire Parameters for Inaudible Single Japanese Vowel Recognition using Myoelectric Signals around a Mouth

永井 秀利 † 宇土 由紀 † 中村 貞吾 †
Hidetoshi Nagai Yuki Uto Teigo Nakamura

1 はじめに

我々は，声を出さずに発声 (いわゆる口パク) された内容を口裂周辺から頸部にかけての位置で体表から測定可能な筋電情報に基づいて認識することを目指している．我々はこれを筋電による黙声認識と呼ぶ．本技術は可聴音を必要としないことにより，他者に騒音で迷惑をかけず盗み聞きもされない (従ってセキュリティ性も高い) 音声入力や大音量下等のマイク入力困難時の音声認識の支援に役立つだけでなく，発声と黙声との切替えにより対話と操作とをシームレスにした音声インターフェース，喉頭切除で声を失った人の発声代行など，多数の応用が考えられる．

我々の従来の研究 [1] では，口裂周辺の 4 筋を計測対象とした日本語母音認識実験において，無理に大きく動かすことをしないような自然な口の動きに対して，発声開始から 100ms ずつの 3 区間の筋電情報を用いて約 87%の認識精度を得た．

動作過程にあり安定性の低い発声開始時の筋電情報を，我々が敢えて用いているのには理由がある．

母音の発声は定常性を持つため，他の黙声 (無発声) 母音認識の研究 ([2] など) では定常動作時の筋電波形を活用するものが多い．しかし，従来の研究 [3] の「あいうえお」連続発声の例にも見られるように，連続発声の場合には定常動作と言える区間が短く，口唇形状の変化動作の中で発声がなされる．それゆえ我々は，黙声単母音の認識においても発声開始時の動作を捉えるような認識を目指している．明確に規定している訳ではないが，「発声開始から 100ms ずつの 3 区間」はそれぞれおよそ発声準備動作，発声開始動作，安定 (定常) 状態を意識している．

日本語母音は口唇形状に特徴があるため口裂周辺の筋によってかなりのレベルでの認識が可能と言える．しかし子音の場合には，口唇形状の変化に加えて舌位置も重要な要素となるため，母音認識を目的とした筋のみでは十分ではない．とはいえ，口腔内にある舌の位置や形状を左右するすべての筋の電位を表面筋電によって計測することはほぼ不可能である．従来の研究 [4] にて，舌根の挙上・下制に関わる筋の一部を計測して子音認識に役立てる試みも行っている[1] が，まだ子音特徴を明確に得るには至っていない．

しかしながら少数単語世界であれば，現在の日本語母音認識手法を援用することにより，ある程度は満足できる精度で認識できる可能性がある．そこで従来の研究 [8] では，それ以前の結果を踏まえ，特性の異なる 2 種の少数語彙世界において，母音認識の場合と同様の筋電信号を用いて単語認識を試みた．

実験の結果として，母音系列としての特徴差が大きい場合には単母音の場合と同じ信号切り出しの窓幅で同等程度の認識精度が得られるのに対し，子音の特徴差を捉えることが必要な少数単語世界で子音特徴の反映を目的に信号切り出しの窓幅を小さくした場合には母音系列としての特徴差を持つ単語間ですら識別精度を大きく低下させてしまうことを確認した．これは，窓幅が小さ過ぎて局所的な変動に過敏となったためであろう．このことは，連続黙声認識を行う際に，短い区間の情報の積み重ねでは十分な認識精度が得られない可能性を示唆する．

そこで本稿では，黙声単母音に対して窓幅や特徴化手法を変更しての認識実験を行った結果に基づいて，筋電信号からの認識パラメータ獲得手法の違いが認識精度に及ぼす影響や特性の違いについて述べる．

2 筋電計測位置

音声認識の場合とは異なり，筋電による認識においては複数の信号を計測し，取り扱う必要がある．

発声時の口唇の形状形成は，複数の筋肉による協調動作の結果である．各筋肉は牽引する方向でしか働かないため，特定の動作を行うには一般に相反する役割を持つ複数の筋肉が係わる．例えば，ある筋肉が同程

†九州工業大学, Kyushu Institute of Technology

[1] 他の研究 [5],[6],[7] でも同様の筋を計測対象に含めている．

表 1: 口筋の種類と作用

分類			筋名	作用
閉口筋			口輪筋	口唇の縮小，収縮，突出
開口筋	上方	浅層	大頬骨筋	口角を外上方に引く
			小頬骨筋	上唇を外上方に引く
			上唇挙筋	上唇，鼻翼ならびに鼻唇溝を上方にあげる
			上唇鼻翼挙筋	上唇を挙上し，鼻孔を拡大する
		深層	口角挙筋	口角を上方にあげる
	外方	浅層	笑筋	口角を後方に引く (頬の皮膚にえくぼを作ることがある)
		深層	頬筋	口の開閉に応じて頬に一定の緊張を与え，咀嚼の際の補助をする また，口角を外後方に引き，口裂を一直線とし，口腔前庭を小さくする
	下方	浅層	口角下制筋	口角を下方に引く
		深層	下唇下制筋	下唇を下方に引く
			オトガイ筋	オトガイの皮膚を上にあげる
			オトガイ横筋	(オトガイの直下を横走する筋肉で，左右の口角下制筋前縁部の筋束が延長してできたもの)

度に働いたとしても，相反する筋肉がどの程度働いたかによって，結果として生じる動作は全く異なるものとなる．そのため，個々の筋肉の信号の強弱だけでは動作状態を正確には判別できず，複数の筋肉の信号の相対関係を調べねばならない．

口唇形状の形成に作用する口筋の種類を表 1 に，これらの内，表面筋電での測定がしやすい浅層に存在する筋のおよその位置を図 1 に示す (ただし，筋肉の幅は再現していない)．これらに加えて，顎の開閉に作用する筋も発声に影響する．なお下唇下制筋は，表 1 のように深層に分類されるているものであるが，同筋の前半部は浅層に存在しているために図に含めている．

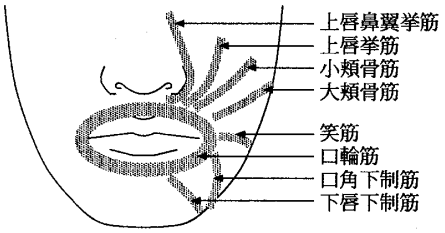

図 1: 浅層の口筋のおよその位置

こうした筋の中からどの筋を対象としてどの位置で計測するかは，重要な問題の一つである．認識に活用できる情報としては多い方が嬉しいが，計測位置が増えるということは利用者の負担の増大や利便性の低下に直結するため，できるだけ必要十分に絞りこむことが望ましい．

口裂周辺の筋電情報に基づいた日本語母音認識を行う他の研究 [9], [10] においては，口筋の中から口輪筋，大頬骨筋の 2 筋に，舌骨挙上または下顎骨引き下げに機能する顎二腹筋を加えた計 3 筋を計測対象としている．大頬骨筋の計測は「い」や「え」の発声において口角を後方に引く動作を捉えようとしたものと考えることができるが，我々の従来の研究では，かなり大きく口を動かさない限りは大頬骨筋に有効な筋電波形を観測することができなかった．角田らの研究 [9] や真鍋らの研究 [10] において，被験者に口唇形状をはっきりさせる発声法の訓練を要求している点からもこの問題[2] の存在は明らかと言えよう．よって日常的な自然な口の動きに対しての認識では大頬骨筋は適切とは言えず，我々は口輪筋，口角下制筋，下唇下制筋[3]，顎二腹筋の 4 筋を計測対象とすることを提案した [1].

表 2: 計測対象とする筋およびその機能

筋名	機能
口輪筋	口唇の縮小，収縮，突出
口角下制筋	口角を外下方に引く
下唇下制筋	下唇の引き下げ
顎二腹筋	舌骨挙上または下顎骨引き下げ

なお，英語を対象とした関連研究では，上唇の形状形成に機能する筋を含めた 7 箇所ないし 6 箇所を計測するもの [5],[6] や，口裂周辺は用いずに下顎から頸部の筋のみを対象とするもの [7] などが存在する．日本語と英語による違いというのもあるが，我々は，日本語において日常的に軽く発声する場合には上唇挙上のような動作はさほど顕著ではないと考えており，前者で挙げられた計測位置の一部は重視していない．後者の場合，頸部のみでは体表から計測できる情報は限られており，少数語彙世界を越えて適用することは非常

[2]もちろん，訓練によって認識に都合が良い発声を行い，精度を向上させることは十分に意義のあることである．だが試してみるとわかるが，それを長時間安定して続けるのはかなりの苦痛と困難とを伴うため，あまり実用的とは言えない．

[3]村木らの研究 [12] では母音ではなく子音判別を目的として下唇下制筋を計測している．

に困難であろうと考える．ただし，従来の研究 [4] でも試みたように，頸部から得られる可能性がある情報には認識の際に重要となりうるものが含まれるため，より多チャネルの計測機器を利用できるようになれば計測対象に含みたい位置である．

3 筋電波形の計測と前処理

本研究では，筋電波形の獲得に 4 チャネルの生体計測器を使用し，図 2 の位置に Ag-AgCl 皿電極を貼り付

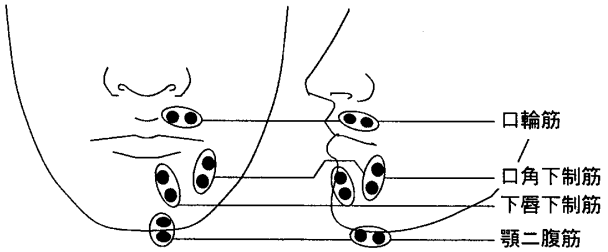

図 2: 表面筋電計測用電極の貼付位置の概略

けて計測を行う．双極誘導による計測のため，各チャネル 2 個，計 8 個の電極を用いる．

電極は，装着位置の皮膚をアルコールで清拭した後，導電性ペーストを用いて装着する．計測は，軽く口唇を閉じた状態 (安静状態) で筋電信号が低く安定した状態から始め，1 母音を発声した後に軽く口唇を閉じた状態に戻すという過程で行う．筋電波形データは，解像度 12bit，周期 50μs(20000Hz) でサンプリングして 1 母音発声 (発声過程の測定時間 2 秒間) ごとに獲得する．フィルタによる帯域制限は 10〜10000Hz した．

筋電信号は微弱であり，獲得した生のデータは多くのノイズを含むため，ノイズをいかに低減するかは重要である．測定対象以外の筋肉からの信号の混入もノイズであるため，周波数帯域の制限では十分なノイズの低減はできない．そこで，ウェーブレット縮退を利用したノイズ低減手法 [3] を適用[4] する．

ウェーブレット縮退においては，どのようなマザーウェーブレットを用いるかによって結果が影響を受ける．我々は，マザーウェーブレットとして Daubechies(N=2) (図 3) を用いている．

ウェーブレット縮退においては，重畳された個々の波形がマザーウェーブレットの形状に近いほど，波形の特徴や細かい変化をうまく拾い上げることができる．

[4]電極の揺れによって生じたと思われるドリフトノイズは取り除くようにしているが，その波形と口の動きとの間には相関性があるように見受けられる．この傾向を調査し有効活用を目指すことも今後の課題の一つである．

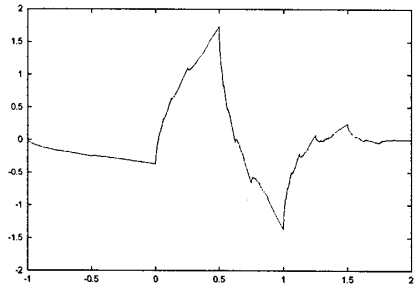

図 3: Daubechie(N=2) のマザーウェーブレット

筋電波形が神経パルスによって生じた信号の重畳した波形であると考えるなら，有界 (コンパクトサポート) かつサポートが狭く，パルスに近い形状を持ったマザーウェーブレットを用いるのが望ましいと言うことができる．Daubechies(N=2) のマザーウェーブレットは，周波数分離性能は良くないが，我々が望ましいと考える特徴を持つ．

本研究において，一般的な筋電信号処理に比べてかなり高いサンプリング周波数を用いているのは，ノイズ低減手法がより良く働くことを期待してのことである．他の筋からの信号 (パルス的信号の重畳) の混入によるノイズの成分をより細かく捉えることで，より良いノイズ低減効果を得ることを目指している．

実際に観測された筋電波形の例と，それに対してノイズ低減処理を行った後の波形とを図 4 に示す．

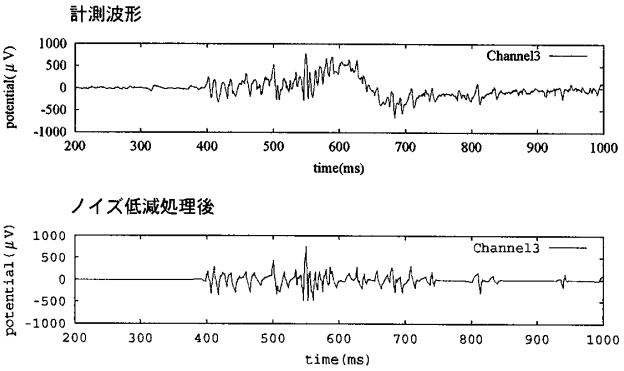

図 4: ノイズ低減処理前後の波形

皮膚との接触抵抗や微妙な位置ずれなどを要因として，電極装着の度に各チャネルの感度には差が生じる．先に述べた通り，筋肉の活動の結果は各筋肉の活動の相対関係によるものであるため，このチャネル間感度差に関する正規化が必要である．また，発声ごとの強度差についても正規化を要する[5]．認識のための波形

[5]従来の研究 [13] で述べたように，発声の強度と筋電信号の強さ

データとしては，これら 2 種類に対しての正規化手法 [1] を適用したものを用いる．

4 認識パラメータの獲得

4.1 波形データの切り出し

前述のように，我々は発声開始時に着目して認識を行おうとしているため，発声開始時の特定が必要ある．従来の研究 [13] や真鍋らの研究 [10] での分析結果でも示されたように可聴音発声時と黙声時とでの筋電波形は十分に類似しているため，可聴音発声時を黙声時の代用とし，発声をマイクで拾うことで発声開始時を定める方法もある．しかし我々は，ノイズ低減後の波形データ (図 4 のすように安静状態の信号は 0 にまで減衰される) における各チャネルの電位が設定した閾値を越えた時点を発声開始時と推定し，その位置を基準として認識に用いる区間を切り出すこととする．これは，可聴音のない現実の黙声認識を想定していると同時に，可聴音が現れる前の準備動作を含めての認識を目指すためである．

我々の従来の研究では，発声開始時と推定された位置から窓幅 100ms ずつで 3 区間，全体で 300ms の範囲の筋電波形を切り出して，認識時のパラメータ生成に利用した．

本稿では，この窓幅の違いが認識精度にもたらす影響を調べる．今回は窓幅を 20ms,40ms,50ms,75ms,100ms の 5 種類とした．切り出し方については，窓の移動幅を窓幅と同じにして窓の重なりがないようにしたものと移動幅を窓幅の半分として重なりを持たせるようにしたものとの 2 種類とした．

切り出す全体幅は，基本として従来と同じ 300ms とした (窓幅 40ms で重なりなしの場合のみ 320ms)．全体幅を揃えるのは，情報源となる筋電波形を同じとすることで，窓幅の差による影響を明確にするためである．これにより，チャネルごとの切り出し区間数は表 3 のようになる．

表 3: チャネルごとの切り出し区間数

窓幅 (ms)	20	40	50	75	100
窓の重なりなし	15	8	6	4	3
窓の重なりあり	29	14	11	7	5

4.2 区間データの特徴量抽出

従来の研究では，筋電波形データから切り出した区間の特徴量の抽出を行う際に計測値の絶対値の総和を用いていた．区間幅は固定されているので，これは ARV (Averaged Rectified Value; 平均整流筋電位) により筋活動電位の量を測っていることに等しい．区間の離散データ $w = \{x_1, \cdots, x_N\}$ に対しての ARV 値 $P_{ARV}(w)$ を式 (1) に示す．

$$P_{ARV}(w) = \frac{1}{N} \sum_i | x_i | \tag{1}$$

ARV は区間全体での平均的な筋活動量を評価するものである．従来の研究では，局所変動やパルス的な信号による悪影響の抑制を目指して ARV を特徴量としたが，筋動作の変遷の速度と比較して区間が広すぎると変化を均してしまうとか，逆に区間が狭いと認識に際して局所的な変化に過敏になる (従来の研究 [8] の実験結果もその一例)．といった問題も存在する．

ARV と同様に，筋活動電位の量的側面を表す特徴量としてよく用いられているものに RMS(Root Mean Square; 二乗平均平方根) がある．区間の離散データ w に対しての RMS 値 $P_{RMS}(w)$ を式 (2) に示す．

$$P_{RMS}(w) = \sqrt{\frac{1}{N} \sum_i {x_i}^2} \tag{2}$$

筋活動水準を測る上では，ARV と RMS とのどちらを用いても大きな違いはないとも言われる．しかしながら，それは骨格筋の運動のような強く継続的な活動の場合に言えることであろう．黙声認識で対象とする筋とその活動のように短い時間で変動する活動では，両者に十分に違いが存在する．

ARV 値と較べると，RMS 値は区間内の強い信号の影響をより受けやすい．そこで，RMS に対してパルス的信号への感度を低下させるため，データ幅 3 での軽い平滑化を加えた後に RMS 値を求める方法も試みた．以下，これを平滑化 RMS (sRMS) と呼ぶ．区間の離散データ w に対しての平滑化 RMS 値 $P_{sRMS}(w)$ を式 (3) に示す．

$$P_{sRMS}(w) = \sqrt{\frac{1}{N} \sum_i \left(\frac{x_{i-1} + x_i + x_{i+1}}{3} \right)^2} \tag{3}$$

ただし x_0 および x_{N+1} は，それぞれ区間内データ x_1 および x_N に隣接する区間外データである．

Jou らの研究 [11] のように特徴化に際してより広い時区間での平滑化を行っている例もあるが，ここでの平滑化の目的は Jou らの研究とは少し異なり，瞬間的

との間には相関が見受けられるため，この正規化を行うことは発声の特徴を一部捨てていることになる．

な強い信号のピークを少しだけ均してやることにある．その意味でデータ幅 3 は十分なものである．本稿では特に実験結果は示さないが，平滑化のデータ幅を 11 にするなど時区間を広げると認識率が低下する傾向が見られた．

平滑化の区間幅を広げた場合の傾向は ARV の結果との比較で捉えることができると考え，本稿ではこれら 3 種を比較する．

他にも，何らかの閾値を用いて筋の活動を 2 値化あるいは少数の段階への離散値化することによって捉えることも可能ではある．しかしながら，前述したように複数の筋の活動の相対関係が重要である場合には，個々の筋の活動の有無が重要となる場合とは異なり，2 値化あるいは少数の段階への離散値化は筋活動の相対関係を捉えることを難しくするだけである．2 値化を行った場合の識別能力の不足などの問題点は真鍋らの研究 [10] によっても確認されており，我々も特徴量の離散値化は考えないこととする．

5 黙声単母音の認識実験

5.1 認識実験方法

本稿の目的は，窓幅や特徴量の違いによる認識率の違いを見ることにある．そのため，認識率に影響を及ぼす他の要因ができるだけ少なくなるように，一人の被験者で，ほぼ休憩なしに，できるだけ短い時間で一度に収集した筋電波形データを実験用データとした．

被験者は一人であるため，個人差の影響が生じることはない．また，計測用の電極を付け替えることなく，皮膚状態の変化も少なくなるように短い時間での計測であることから，チャネル間感度差の正規化が及ぼす影響を心配する必要もない．ノイズ低減処理を行うことは避けられないが，それ以外には発声ごとの強弱の正規化のみを施すだけで良いことになる..

その代償として，サンプルとして収集できる筋電波形データの数が少なくなり，窓幅などについて絶対的な数値としての良否の断言はできなくなるが，我々はそれでも認識パラメータ獲得方法による傾向の違いを見ることはできると考えた．

計測は 3 章で述べた方法で行い，ノイズ低減処理と正規化とを施した結果として，各日本語母音につき 19 個，総計 95 個の筋電波形データを得た．これらの波形データに対し，4 章で述べた窓幅，窓移動幅，特徴量によって認識用パラメータを獲得した．

認識実験には，ごく一般的な 3 階層のフィードフォワード型ニューラルネットワークを用いた．各日本語母音 1 個ずつ，計 5 個のデータを一つのまとまりとして 19 個のデータセットを作り，内一つをテストデータ，残りを学習データとして計 19 回の認識実験を行い，認識精度を求める．これを，窓幅，窓移動幅，特徴量の組み合わせ (全 30 パターン) のそれぞれについて行った．

5.2 実験結果と考察

認識実験の結果を表 4 から表 9 までに示す．表に示された通り，テストデータに対する認識率が最大となったのは，窓幅 75ms，窓の重なりあり，平滑化 RMS の場合で，認識率は 95.7%であった．

各特徴量での最大認識率を比較してみると，窓の重なりなしの場合で ARV が 86.3%，RMS が 92.6%，平滑化 RMS が 94.7%，窓の重なりありの場合で ARV が 88.4%，RMS が 93.6%，平滑化 RMS が 95.7%であった．窓の重なりの有無のそれぞれにおいて，いずれの場合も平滑化 RMS がもっとも高い認識率を得ることができた．

窓幅と特徴量とを固定した場合，いずれの場合でも窓の移動幅を窓幅の半分にして重なりを持たせた場合の方が高い認識率を示した．

重なりの有無と特徴量とを固定した場合，認識率はいずれの場合でも窓幅 75ms の時をピークとし，これから離れる程に低下する傾向にある．この 75ms という値は，真鍋らの研究 [10] で述べられている「フレーム長が 80ms 以上であれば，筋電信剛の時間的な変動が認識結果に影響を与えないと解釈できる」とする主張とも整合する．真鍋ら [10] の主張とは異なり 100ms の場合に 75ms の場合よりも認識率が低下してしまう理由は，本研究が発声開始時の動作を重視している点にあると考える．発声時の口唇形状形成動作の速度やリズムに対して 75ms という窓幅が実験した中では最も適合しており，これよりも長くても短くても窓幅の適合性が悪い (発声準備動作と発声動作との変化点を中途半端に跨いでしまうことなど) ことによる悪影響が増大して認識率の低下が生じているものと考える．

以上のように多くの場合で平滑化 RMS が優れるが，窓幅と重なりの有無とを固定して見た場合には RMS と平滑化 RMS とで順位が逆転するケースがある．その様子を見るため，誤認識率によってグラフ化したものを図 5 に示す．

グラフから，窓の重なりの有無によらず，各特徴量

表 4: ARV を用いた窓重なりなしの認識率 (%)

窓幅	100ms		75ms		50ms		40ms		20ms	
母音	学習	テスト	学習	テスト	学習	テスト	学習	テスト	学習	テスト
あ	89.7	68.4	94.4	78.9	97.0	73.6	93.5	78.9	94.7	63.1
い	100.0	94.7	99.7	89.4	99.1	89.4	99.1	89.4	97.3	78.9
う	97.9	89.4	97.9	89.4	99.1	89.4	97.9	84.2	97.3	89.4
え	90.0	78.9	96.4	89.4	98.2	84.2	97.3	84.2	96.1	78.9
お	99.1	89.4	99.4	84.2	97.3	89.4	98.8	84.2	98.8	89.4
全体	95.1	84.2	97.3	86.3	98.1	85.2	97.3	84.2	96.9	80.0

表 5: ARV を用いた窓重なりありの認識率 (%)

窓幅	100ms		75ms		50ms		40ms		20ms	
母音	学習	テスト	学習	テスト	学習	テスト	学習	テスト	学習	テスト
あ	94.7	73.6	93.5	78.9	99.1	84.2	95.3	73.6	92.6	63.1
い	100.0	94.7	99.4	89.4	99.1	94.7	98.8	89.4	99.7	89.4
う	99.7	89.4	99.4	94.7	99.7	84.2	99.1	94.7	94.1	89.4
え	95.3	84.2	97.6	84.2	100.0	89.4	97.6	84.2	89.4	73.6
お	97.9	89.4	99.7	94.7	100.0	89.4	99.1	89.4	100.0	89.4
全体	97.6	86.3	97.9	88.4	99.5	88.4	98.0	86.3	95.4	81.0

表 6: RMS を用いた窓重なりなしの認識率 (%)

窓幅	100ms		75ms		50ms		40ms		20ms	
母音	学習	テスト	学習	テスト	学習	テスト	学習	テスト	学習	テスト
あ	100.0	94.7	99.7	89.4	99.7	89.4	100.0	89.4	99.4	89.4
い	100.0	94.7	100.0	94.7	100.0	94.7	100.0	94.7	99.4	89.4
う	100.0	94.7	99.4	94.7	100.0	94.7	100.0	94.7	100.0	94.7
え	99.1	78.9	99.1	89.4	99.1	78.9	99.7	84.2	100.0	78.9
お	100.0	94.7	100.0	94.7	100.0	100.0	100.0	89.4	100.0	89.4
全体	99.8	91.5	99.6	92.6	99.7	91.5	99.9	90.5	99.7	88.4

表 7: RMS を用いた窓重なりありの認識率 (%)

窓幅	100ms		75ms		50ms		40ms		20ms	
母音	学習	テスト	学習	テスト	学習	テスト	学習	テスト	学習	テスト
あ	99.7	89.4	99.1	89.4	99.1	84.2	99.1	89.4	99.1	84.2
い	100.0	94.7	100.0	94.7	100.0	94.7	100.0	94.7	98.5	89.4
う	99.1	94.7	98.8	94.7	99.4	94.7	99.7	94.7	100.0	94.7
え	97.3	89.4	97.0	89.4	97.9	89.4	98.2	84.2	97.3	78.9
お	100.0	94.7	100.0	100.0	100.0	100.0	99.4	94.7	100.0	94.7
全体	99.2	92.6	99.0	93.6	99.2	92.6	99.2	91.5	99.0	88.4

表 8: 平滑化 RMS を用いた窓重なりなしの認識率 (%)

窓幅	100ms		75ms		50ms		40ms		20ms	
母音	学習	テスト	学習	テスト	学習	テスト	学習	テスト	学習	テスト
あ	99.4	94.7	99.7	94.7	94.1	89.4	97.3	89.4	98.2	84.2
い	100.0	94.7	100.0	94.7	98.2	94.7	100.0	89.4	100.0	89.4
う	97.0	94.7	100.0	100.0	100.0	89.4	99.7	94.7	99.7	89.4
え	94.1	84.2	95.0	84.2	97.6	84.2	100.0	84.2	99.1	84.2
お	99.7	89.4	100.0	100.0	100.0	100.0	100.0	84.2	100.0	84.2
全体	98.0	91.5	98.9	94.7	98.0	91.5	99.4	88.4	99.4	86.3

表 9: 平滑化 RMS を用いた窓重なりありの認識率 (%)

窓幅	100ms		75ms		50ms		40ms		20ms	
母音	学習	テスト	学習	テスト	学習	テスト	学習	テスト	学習	テスト
あ	100.0	89.4	100.0	100.0	97.0	89.4	98.5	84.2	99.1	84.2
い	100.0	94.7	100.0	94.7	100.0	89.4	100.0	89.4	100.0	84.2
う	97.0	94.7	99.7	94.7	98.8	94.7	99.1	94.7	99.1	94.7
え	97.3	89.4	95.3	89.4	100.0	84.2	99.7	84.2	97.9	84.2
お	100.0	94.7	100.0	100.0	100.0	100.0	99.7	89.4	100.0	89.4
全体	98.8	92.6	99.0	95.7	99.1	91.5	99.4	88.4	99.3	87.3

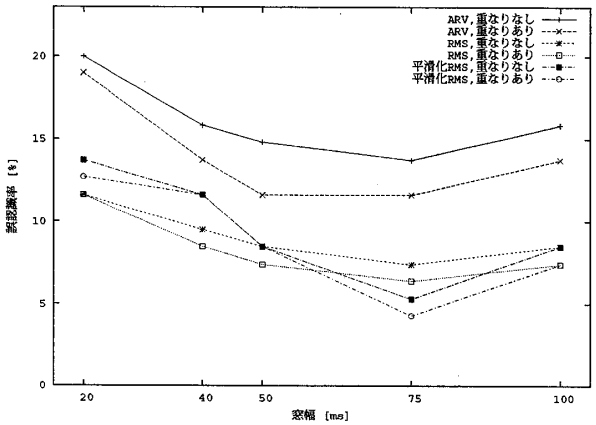

図 5: 特徴量ごとの誤認識率

ごとの傾向が存在することが見て取れる．

平滑化 RMS を用いた場合，最適な窓幅時には最良の結果を得ることができるが，その窓幅から離れるにつれて認識率が低下していく度合は他の特徴量よりも大きい．通常の RMS の方が認識率低下の度合が緩やかであるため，最適時以外では通常の RMS と同等以下の認識率となってしまっている．

本稿で取り扱う範囲を越えるため今後の研究課題とはなるが，もし最適な窓幅が発声時動作のタイミングとの適合性に依存するのであれば，発声速度の変動への耐性が問題となる可能性がある．その場合には，発声速度に動的に対応できるようにする技術を目指すか，次善策として通常の RMS で妥協するかの選択が必要となるかもしれない．もちろん，安定した速度で発声するようにすることで，平滑化 RMS による最良の認識率を活かすという選択肢も有効であろう．

次に，母音ごとの認識状況を見る．

従来の黙声母音認識 [1] で用いていたものに相当する ARV，窓幅 100ms，重なりなしでのテストデータに対する認識結果を表 10 に，少数単語世界での認識 [8] において孤立数字認識に試用したものに相当する ARV，窓幅 40ms，重なりありでの結果を表 11 に示す．また，本稿での実験で最も認識精度が良かった窓幅である窓幅 75ms 時の各設定でのテストデータに対する認識結果を表 12 から表 17 に示す．

従来の研究で用いていた手法 (表 [10] や表 [11]) では，「あ」と「い」，「あ」と「え」，「い」と「え」，「う」と「お」といった間での誤認識が多く見られる．これらは，自然な発声を行った場合に軽い動作となり活動差が微弱となりやすい顎二腹筋 (顎の開閉の判別に影響) や口角下制筋 (口角の固定や外方への誘引) を特徴差の一部とする組み合わせである．口輪筋が特徴差の一部となるような組み合わせについては誤認識は少なく，従来の手法でも口輪筋については十分に特徴を捉えていると言える．認識精度が最も高かった窓幅 75ms の時 (表 [12] や表 [13]) と比較しても，特徴量として ARV を用いた場合の誤認識傾向には違いがない．

軽い動作の場合は，口唇形状形成時に短時間だけやや強い活動が生じ，発声開始後は脱力か，あるいは形状維持程度の弱い活動へと移行する．例えば「あ」の発声は，発声準備として顎を開く際の短時間だけ顎二腹筋に強めの活動が生じ，その後は重力によって顎が下がるに任せて顎二腹筋は脱力するというような傾向が見られる．本稿での実験結果は，特徴量として ARV を用いたのではこのような特徴をうまく捉えることが難しいことを示している．

RMS および平滑化 RMS を用いた場合には，ARV において誤認が多かった組み合わせ (我々の従来の研究を含む) に対してかなり改善されている．特に「あ」の認識において，改善が顕著である．認識結果がやや「あ」に引っ張られ気味であり，「あ」の改善もそれに依存している可能性はあるが，全体として ARV よりも改善されているという傾向は明らかであり，軽い動作での特徴をより良く捉えていると考える．

6 おわりに

連続黙声母音認識に向けて，黙声単母音の発声開始時の動作を捉える形での黙声単母音認識を行った．その際，いくつかの認識パラメータ獲得方法を比較し，窓幅 75ms，窓移動幅を窓幅の半分として筋電波形を切り出し，平滑化 RMS を特徴量として用いることで，黙声単母音を 95.7%の精度で認識することができた．窓移動幅を変更した場合も含め，発声速度の変化への追随性の調査，分析は今後の課題ではあるものの，特徴量として平滑化 RMS が良い認識精度を与えることを示すと同時に，窓幅選択の一つの基準として 75ms という値を示した．

今後は，話者への依存性の調査，分析を行って本稿で得られた結果の精緻化を進めると同時に，本稿での結果を踏まえて連続黙声母音認識へと研究を発展させることが課題である．

参考文献

[1] 永井, 中山, 中村, 野村: “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”, 電気関係学会九州支部大会 12-1P-05 (2004)

表 10: ARV, 窓幅 100ms, 重なりなしの認識結果

発声＼認識	あ	い	う	え	お
あ	13	3	0	3	0
い	0	18	0	1	0
う	0	0	17	0	2
え	2	2	0	15	0
お	0	0	2	0	17
合計	15	23	19	19	19

表 11: ARV, 窓幅 40ms, 重なりありの認識結果

発声＼認識	あ	い	う	え	お
あ	14	1	0	4	0
い	0	17	0	2	0
う	0	0	18	0	1
え	0	3	0	16	0
お	0	0	2	0	17
合計	14	21	20	22	18

表 12: ARV, 窓幅 75ms, 重なりなしの認識結果

発声＼認識	あ	い	う	え	お
あ	15	3	0	1	0
い	0	17	0	1	1
う	0	0	17	0	2
え	1	1	0	17	0
お	0	1	2	0	16
合計	16	22	19	19	19

表 13: ARV, 窓幅 75ms, 重なりありの認識結果

発声＼認識	あ	い	う	え	お
あ	15	3	0	1	0
い	0	17	0	2	0
う	0	0	18	0	1
え	1	2	0	16	0
お	0	0	1	0	18
合計	16	22	19	19	19

表 14: RMS, 窓幅 75ms, 重なりなしの認識結果

発声＼認識	あ	い	う	え	お
あ	17	1	0	1	0
い	1	18	0	0	0
う	0	0	18	0	1
え	2	0	0	17	0
お	0	0	1	0	18
合計	20	19	19	18	19

表 15: RMS, 窓幅 75ms, 重なりありの認識結果

発声＼認識	あ	い	う	え	お
あ	17	0	0	2	0
い	1	18	0	0	0
う	0	0	18	0	1
え	2	0	0	17	0
お	0	0	0	0	19
合計	20	18	18	19	20

表 16: 平滑化 RMS, 窓幅 75ms, 重なりなしの認識結果

発声＼認識	あ	い	う	え	お
あ	18	0	0	1	0
い	1	18	0	0	0
う	0	0	19	0	0
え	3	0	0	16	0
お	0	0	0	0	19
合計	22	18	19	17	19

表 17: 平滑化 RMS, 窓幅 75ms, 重なりありの認識結果

発声＼認識	あ	い	う	え	お
あ	19	0	0	0	0
い	1	18	0	0	0
う	0	0	18	0	1
え	2	0	0	17	0
お	0	0	0	0	19
合計	22	18	18	17	20

[2] 張, 真鍋, 平岩, 杉村: “HMM 及びケプストラム係数特徴による筋電信号を用いた無発声音声認識”, 電子情報通信学会技術報告 Vol.103, No.401, pp.7–12(2003)

[3] 永井, 中村, 野村: “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「火の国シンポジウム 2003」, pp.1–8(2003)

[4] 永井, 南, 中村, 野村: “筋電に基づく黙声認識における子音認識のための基礎的調査”, 電気関係学会九州支部大会 12-1P-06 (2004)

[5] Maier-Hein, Metze, Schultz, and Waibel: “Session Independent Non-Audible Speech Recognition Using Surface Electromyography”, Proc. ASRU, pp.331–336(2005)

[6] Jou, Schultz, and Waibel: “Continuous Electromyographic Speech Recognition with a Multi-Stream Decoding Architecture”, Proc. ICASSP, Vol.4, pp.401–404(2007)

[7] Betts and Jorgensen: “Small Vocabulary Recognition Using Surface Electromyography in an Acoustically Harsh Environment”, tech. memo TM-2005-213471, NASA (2005)

[8] 永井, 谷口, 副島, 中村, 野村: “口裂周辺の筋電信号を用いた少数語彙世界における黙声単語認識”, 第 6 回情報科学技術フォーラム (FIT2007) 論文集,E-023(2007)

[9] 角田, 杉江: “音声合成方式発声代行システム —筋電位信号からの母音の判別と発声—”, 電気学会論文誌 105-C, pp.25–32(1985)

[10] 真鍋, 平岩, 杉村: “無発声音声認識 : 筋電信号を用いた声を伴わない日本語 5 母音の認識”, 電子情報通信学会論文誌 D-II, Vol.J88-D-II, No.9, pp.1909–1917(2005)

[11] Jou,Schultz,Walliczek,Kraft, and Waibel: “Towards Continuous Speech Recognition Using Surface Electromyography”, Interspeech 2006, pp.573–576(2006)

[12] 村木, 角田, 杉江: “代用発声のための子音判別法に関する基礎的研究”, 電子通信学会技術報告, MBE83-108(1983)

[13] 永井, 中村, 野村: “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)