

口部パターン形状を利用した文字入力システム

齊藤 剛史[†] 加藤 友哉[†] 小西 亮介[†]

[†] 鳥取大学工学部, 〒 680-8552 鳥取市湖山町南 4-101

E-mail: †{saitoh,konishi}@ele.tottori-u.ac.jp

あらまし 本論文では、口部形状のリアルタイム認識と簡単な首の動きを利用した、従来にない新しい文字入力システムを提案する。文字の入力は、ユーザが日本語の五十音における入力したい文字の属する段の母音形状を作る。母音認識の後にその段の文字を走査選択式により順次自動的に選び、ユーザの頷き動作により文字が入力される。7~10名の被験者の協力を得て母音形状認識実験、首振り動作実験および文章入力実験を行った。その結果、5文の文字入力速度は1分間に平均6.8文字であり、入力ミス回数は4.3%であった。キー操作を利用する方法などに比べ入力速度は速くないが、人に自然なコミュニケーション手段である発話を意識した口の動きによる文字入力を実現した。

キーワード 口部パターン, 読唇, 母音認識, テキスト入力システム

Characters Entry System based on Oral Pattern Shape

Takeshi SAITOH[†], Tomoya KATO[†], and Ryosuke KONISHI[†]

[†] Tottori University, Tottori

E-mail: †{saitoh,konishi}@ele.tottori-u.ac.jp

Abstract This paper proposes a novel characters entry system based on oral shape pattern and simple nod or shake motion. To entry a character, user utters a vowel which is same row in the Japanese syllabary. Then, vowel recognition process is carried out, and character list is displayed. When user nods, one of the characters is inputted. Three experiments, vowel recognition, nodding or shaking motion analysis, and text entry experiment, were carried out. As the results, average character entry speed was 6.8 kana per minute, and average miss entry time was 4.3% with five sentences. Though character entry speed of our method is not faster than other methods, we realized the system that conscious the utterance which is natural communication method.

Key words oral shape, lip reading, vowel recognition, text entry system

1. はじめに

コンピュータや携帯電話などの機器の普及により、多くの人がこれらの機器を用いて文章を入力する機会が増えている。文字を入力するための一般的なインタフェースは、キーボードやマウス、ボタンなどの手で操作する手段である。しかし、全ての人がキーボードやマウスなどを巧みに操作して文字を入力できるとは限らず、手指に障害をもつ人はこれらの操作が難しい。このことは高齢化社会を迎えるうえで、次第に重要視される問題であると考えられる。

手指を使わないインタフェースとして、音声認識や視線を利用する方法が提案されている。また、唇の動きを利用して発話内容を認識する読唇に関する研究も取り組

まれている[1]~[4]。しかし、音声認識に比べ認識率が低く、まだ実用段階にあるとはいえない。

本論文では、コンピュータや携帯電話などの機器を操作するのに、手指を用いずに口部形状と簡単な首の動きを利用して文字を入力する手法を提案する。提案手法は、キー操作を利用する方法などに比べ、文字の入力速度は早くないが、人の自然なコミュニケーション手段である発話を意識した口の動きで文字の入力を実現する。

1.1 従来のインタフェース

文字の入力システムとしてこれまで様々なインタフェースが提案されている。

1) コンピュータなどに使用されているキーボードのように、全ての文字がボタンに表示されており、そのボタ

ンを押すことにより文字を入力する方法がある。文字と1体1で対応しているボタンを押すため、操作が簡単で精度も高いが、ボタンを文字の個数分用意する必要があり、手指の細かな動作が必要となる。

2) 携帯電話などで使用されている一つのボタンを複数回押すことにより文字を入力する方法がある。1) に比べて少ないボタン数で済むが、操作回数が増大し、ユーザの操作が煩雑になる。

3) ペンなどを使用して画面などに文字を描き、手書き文字認識技術 [5] を利用して文字を入力する方法がある。この方法は、日常生活で用いられている文字を書く感覚で入力できるため、操作が容易であるが、ペンなどの道具や文字を描くための場所が必要になる。

4) 音声認識技術を利用して、音声を発話して文字として入力する方法がある [6]。発話するだけで文字の入力が可能なため使用が容易であり、3) のように道具を必要としない。しかし、周囲の騒音に影響を受ける問題がある。

5) 視線測定技術による視線を用いたインタフェースが提案されている [7], [8]。視線インタフェースではユーザが視線を利用して能動的に対象を選択する方法と、コンピュータ側が利用者の視線を観察して、観察結果に基づき作業支援を行う手法がある。視線インタフェースでは、意図的に視線をかえるのはユーザに負担を強いる問題がある。また、視線を用いるのは、単に対象物に注視した状態と、装置に対する何らかの意図をもって見ているのかを判別することが難しいなどの問題もある。

6) Ward と MacKay はキーボードを使わずに、マウスやスタイラス（ペン型入力装置）、視線などを用いて、文字を順々に選択していくことにより、文章の入力が行えるインタフェース Dasher を提案している [9]。Dasher では画面上に入力頻度の高い文字が次々に流れていくことで高速入力が可能になる。

7) Lyons らは口の形状とキー操作を併用したテキスト入力システムを提案している [10]。日本語のかな五十音表の異なる子音に対してはキーを、母音に対しては口の形状を対応させることによりテキスト入力を実現している。キー操作を利用することにより早いテキスト入力を実現しているが、テキストを入力するためにキー操作が必要であり、手指の細かな動きが重要である。

1.2 アプローチ

読唇技術に関して、我々は日本語 45 音の単音の発話に対して、5 母音、10 子音、45 音への分類を試みた [11]。その結果、5 母音への分類は約 94%、10 子音と 45 音への分類は共に約 29% の識別率を得た。この結果より、単音への識別精度が低く、ユーザに入力する文字を直接発話するのは有用でないと考えられる。一方、母音への分

類は高い識別精度を得られている。

また、文字の入力方式として直接選択式と走査選択式がある [8]。直接選択式は、一般のキーボードのように、入力したい文字を直接選べる方式である。走査選択式は、選択候補群の中から候補が一定時間ごとに順次自動的に選ばれ、望みの候補が選ばれるときにスイッチをオンすることで入力する方法である。直接選択式に比べ、入力速度は遅くなるが、1~2 個の入力スイッチしか操作できない障害者では、それ以外の入力手段がないとして採用されることが多い入力方式である。

以上のことより本研究では、入力したい文字の単音を発話するのではなく、五十音表における入力したい文字の属する段の母音形状を作る。母音を認識した後に、その段の文字を走査選択式により順次自動的に選び、ユーザが頷く（首を下に向ける）動作をすることによりスイッチをオンにしたと判断し文字を入力する方法を採る。

2. 特徴量の算出

2.1 口内領域の抽出

画像情報に基づく読唇法では特徴量の計測が重要であり、これには画像ベース法とモデルベース法がある。前者は口唇周辺領域の画素濃度値を利用するため、唇形状だけでなく歯や舌の情報を含めた特徴量を取得できる利点がある。しかし、画素情報に基づくため多量のデータが必要であり、領域サイズや撮影環境における明暗の影響を受けやすい問題がある。一方、後者は 2 値化処理や動的輪郭モデルなどにより口唇領域のモデルを計測し、口唇領域の幅や面積などの少ない情報でモデルを表現できる利点がある。本研究では、先行研究 [4] で用いた口内領域を 2 値化処理により得る。領域抽出および領域の正規化の詳細は文献 [4] を参照されたい。

2.2 口部形状認識の特徴量

口部形状の認識には、単音認識 [4] で有効であった面積 S とアスペクト比 $A = H/W$ の 2 特徴量を用いる。ただし、 H 、 W はそれぞれ口内領域の縦幅と横幅である。単音認識 [4] では 2 特徴量の時間的変化であるトラジェクトリ特徴量を用いていた。しかし本論文では、口部形状を認識する際に形状の静止判定を用いる。そこで、トラジェクトリ特徴量を用いずに時間 t の 2 特徴量 $\alpha(t) = (S(t), A(t))$ を用いる。

算出した特徴量を図 1 に示すように視覚的にユーザへ提示する。図中、横軸は面積 S 、縦軸はアスペクト比 A である。また十字の交点が現在の特徴量 $\alpha(t)$ の位置であり、時間的変化をプロット点の明暗で表現している。白いプロット点ほど過去、黒いプロット点ほど現在に近い時間である。交点周辺の円は静止判定に用いるしきい値である。また図には学習用データとして登録した口部形

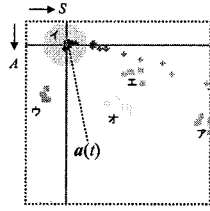


図1 特徴量の推移図

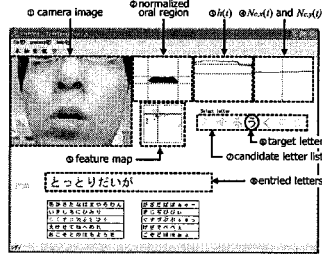


図2 提案システムの画面構成

状を色分けして表示している。

3. 文字入力システム

本論文で提案する、口部形状と首の動きを利用した文字入力システムは形状登録モードと文字入力モードの二つから構成されている。

3.1 システムの画面構成

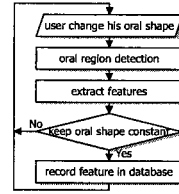
本論文で構築した文字入力システムを図2に示す。画面内には①カメラ画像、②正規化後の口内領域画像、③口内領域の高さの変化 $h(t)$ 、④鼻孔間中点 N_c の x 、 y 座標の変化、⑤計測された2特徴量の推移図、⑥認識結果あるいは入力対象文字、⑦入力候補文字リスト、⑧入力された文字を表示する。また最下部には本システムで入力可能な文字の一覧を表示する。

3.2 形状登録モード

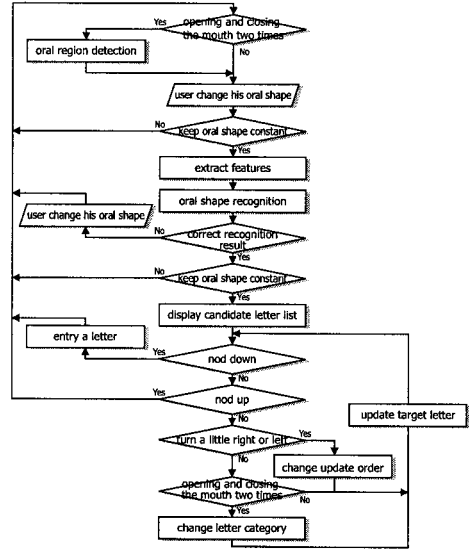
形状登録モードでは、ユーザ毎の口部形状を登録する。処理の流れを図3(a)に示す。

I) ユーザは登録したい口部形状を作る。入力画像より口内領域の抽出処理および口内領域の面積とアスペクト比の2特徴量が計算される。

II) ユーザは一定時間 T_s 同じ口部形状を保つと、現在の口部形状を登録するかメッセージが表示される。ここで本論文では、一定時間同じ口部形状であるかを判断することを静止判定と呼ぶ。静止判定は現在の特徴量 $a(t_0)$ を基準とし、過去 T_s 時間までの特徴量 $a(t_0 - t)$, $0 < t < T_s$ の最大誤差が特徴量空間においてしきい値以下である場合、すなわち $\max |a(t_0) - a(t_0 - t)| < \epsilon$ を満たす場合、静止していると判定する。しきい値 ϵ は図1で円の半径として示されている。



(a) 形状登録モード



(b) 文字入力モード

図3 処理の流れ

3.3 文字入力モード

文字入力モードでは、前節で登録したユーザ毎の口部形状を利用して口部形状の認識を行い、その結果に応じて首の動作を行うことにより文字を入力する。処理の流れを図3(b)に示す。

i) ユーザは入力したい文字の母音形状、例えば‘わ’であれば‘あ’の形状を作り、登録モードと同様に一定時間 T_s 同じ口部形状を保つ。

ii) 静止判定を行い、静止している場合は画面(図2⑥)に認識結果が表示される。そうでない場合はi)に戻る。

iii) 認識結果が正しければiv)へ、間違っていれば口の形状を変えることによりi)に戻る。

iv) 正しい認識結果であれば、さらに一定時間 T_s 同じ形状を保つと、入力文字の候補リストが図2⑦に表示される。例えば、‘あ’の形状であれば、あ段の文字リストが表示される。文字リストは T_u 間隔で自動更新されて順番に表示される。文字リストの更新順序はデフォルトでは昇順であるが、後述の首の左右の動きにより降順にかえることができる。

v) 表示のタイミングを見計らって以下の首の動きをする。ここで口部形状の情報は用いなくなるため、ユーザは口を閉じるなどの自然な状態であり。

○ 文字を入力するとき、入力したい文字が強調表示される直前に、頷くように下を向いて再び顔を正面の位置に戻す動作（下向き動作）を行う。強調文字は図 2⑥に表示される。下向き動作が入力されると図 2⑧に文字が追加される。

○ 文字の入力をキャンセルするとき、文字の入力とは逆に上を向いて再び顔を正面の位置に戻す動作（上向き動作）を行う。上向き動作が入力されると図 2⑦の入力文字の候補リストが表示されなくなる。

○ 候補リストの更新順序を降順にするとき、左を向いて再び顔を正面の位置に戻す動作（左向き動作）を行う。

○ 候補リストの更新順序を昇順にするとき、右を向いて再び顔を正面の位置に戻す動作（右向き動作）を行う。

○ 濁音や半濁音などの文字を入力するとき、口の 2 回の開閉動作（開閉動作）を行う。

vi) 入力した文字を取り消すとき、認識モードでない状態で開閉動作を行うことにより最後に入力した 1 字を消すことができる。開閉動作が入力されると図 2⑨の最右端の入力文字が 1 字消える。

3.4 首振り動作

前述の通り、本論文で提案する文字入力システムでは、下向き、上向き、右向き、左向きの四つの首の動きと口の開閉動作を利用する。各動作について図 4 を用いて説明する。

下向き動作および上向き動作は首の縦方向の動きである。ここでは画像内において首の動きに応じて位置が変化する鼻孔間の中点 N_c に注目する。すなわち、 N_c が下方に移動する場合は下向きの動きが、上方に移動する場合は上向きの動きが生じたと考えられる。そこで N_c の y 座標の時間的変化 $N_{c,y}(t)$ に注目し、下記 2 条件をともに満たすときに下向き動作あるいは上向き動作が入力されたと判断する。

$$(1) |N_{c,y}(t) - N_{c,y}(t - T_m/2)| > Th$$

$$(2) |N_{c,y}(t) - N_{c,y}(t - T_m)| < Th$$

ただし、 T_m は首の動きの動作判定の解析時間であり、 Th は y 座標の動き量に対するしきい値である。第 1 条件で、過去の移動量がしきい値を満たしているか判定する。 $t - T_m/2$ を用いるのは、首の向きが正面から上あるいは下に移動して再び正面の向きに戻るときの移動量、移動時間は往路、復路ともにほぼ同じであり、かつ処理時間の短縮を図るためである。また第 2 条件で、首の向きが元の向きに戻っているか判定する。また $N_{c,y}(t) < N_{c,y}(t - T_m/2)$ のときは下向き、そうでない場合は上向き動作と判定する。

左向き動作および右向き動作は首の横方向の動きである。これは前述の縦方向の動きにおいて $N_{c,y}(t)$ に注目したのに対し、 N_c の x 座標の時間的変化 $N_{c,x}(t)$ に注目する。動作条件は下記の通りである。

$$(3) |N_{c,x}(t) - N_{c,x}(t - T_m/2)| > Th$$

$$(4) |N_{c,x}(t) - N_{c,x}(t - T_m)| < Th$$

$N_{c,x}(t) < N_{c,x}(t - T_m/2)$ のときは右向き、そうでない場合は左向き動作と判定する。

四つの首振り動作の例を図 4(a)~(d) に示す。図 4(a)~(d) のそれぞれの三つの画像は $N_{c,*}(t - T_m)$ 、 $N_{c,*}(t - T_m/2)$ 、 $N_{c,*}(t)$ を示している。ただし、 $*$ は x あるいは y を意味する。左図の破線で囲まれた位置で動作判定を満たしている。

口の開閉動作は、正規化後の口内領域の高さの時間的変化 $h(t)$ において、口の開閉の動作判定の解析時間 T_o 内において、2 条件 $h(t) < Th_l$ 、 $h(t) > Th_u$ を交互に 2 回ずつ満たすときに開閉動作と判定する。口の開閉動作の例を図 4(e) に示す。左図の破線で囲まれた位置で動作判定を満たしている。

4. 実験

4.1 母音形状認識実験

1.2 で述べたとおり、本論文では認識対象として、日本語 5 母音（‘ア’、‘イ’、‘ウ’、‘エ’、‘オ’）の 5 形状とした。10 人の被験者（A~J は全て成人男性健常者、G は留学生、それ以外は日本人）に対して 5 母音形状の認識実験を行った。実験では、被験者は机上に置かれたカメラに向かって椅子に着席させ、自然な母音の発話を意識させた。学習データとして被験者自身の発話した母音 5 形状をそれぞれ 5 個ずつ計 25 個用いた。認識は 5 形状について 20 回ずつ計 100 回実験を行った。ただし本実験では、市販の USB カメラを用いて取得した 160×120 画素の画像に対して処理を適用した。使用した計算機は DOS/V PC (CPU:2.01GHz) であり、処理速度は約 28fps であった。また、口部形状の静止判定の時間は $T_s = 1.0\text{sec}$ とした。被験者とカメラ間の距離は特に指定していないが、およそ 15~30cm の範囲であった。

認識結果を表 1 に示す。被験者により認識率の差はあるものの平均認識率は 77.4% であった。これは文献 [4] の結果に比べ低い認識率である。この原因を調べるため、学習データの特徴量空間の分布を調べた。その結果を図 5 に示す。図中のプロット点は学習データとして登録された 25 個の口部形状の位置、領域の色分けは、各位置で認識される結果を示している。図 5(a) 左は認識率の高い被験者 D、図 5(b) 左は認識率の低い被験者 E の分布である。認識率の低い被験者は図 5(b) 左のように母音形状の分布が混在していたため誤認識が多いと推測する。

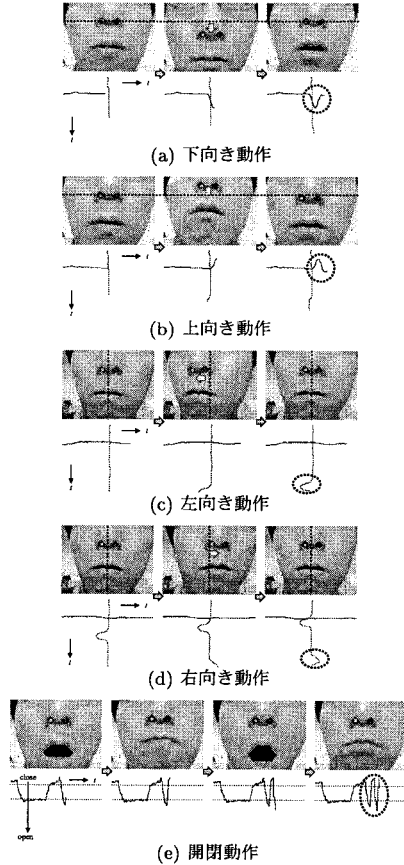


図4 首振り動作と開閉動作例
表1 5母音形状の認識率

vowel	subjects										ave.
	A	B	C	D	E	F	G	H	I	J	
'ア'	100.0	100.0	95.0	100.0	40.0	80.0	75.0	95.0	85.0	80.0	85.0
'イ'	85.0	80.0	95.0	80.0	60.0	100.0	55.0	100.0	45.0	80.0	78.0
'ウ'	100.0	85.0	60.0	100.0	90.0	80.0	85.0	85.0	35.0	0.0	72.0
'エ'	75.0	75.0	100.0	100.0	30.0	70.0	55.0	95.0	75.0	60.0	73.5
'オ'	90.0	75.0	95.0	100.0	70.0	80.0	45.0	95.0	85.0	50.0	78.5
ave.	90.0	83.0	89.0	96.0	58.0	82.0	63.0	94.0	65.0	54.0	77.4

また、Lyonsら[10]はキー入力を利用するため、ユーザが意図するときに認識される。一方、本手法ではキー入力を利用せずに静止判定を適用することにより認識する。すなわち、ユーザが意図する時間 T_u と認識される時間 T_c に差 $\Delta T = T_u - T_c$ が生じる。そこで、本実験では ΔT を計測し、リアルタイム認識の有用性についても検討した。その結果、10人の被験者の平均時間差は0.55 secであった。Lyonsらの手法は $\Delta T = 0$ であるが、本手法も $\Delta T < 1.0$ secであり、ユーザの意図する時間との差が少なく、有用であると判断できる。

4.2 首振り動作実験

本実験は、7名の被験者(前実験のA~G)に対して前実験と同様な姿勢をとらせ、さらに被験者にマーカー付

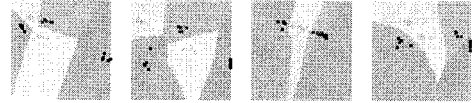


図5 被験者毎の特徴量空間の分布図
表2 首振り動作の実験結果

motion	θ_m [deg]			t_m [sec]		
	ave.	min	max	ave.	min	max
NU	9.61	4.71	14.56	0.78	0.53	1.37
ND	10.42	5.12	18.99	0.74	0.47	1.27
TR	9.34	5.17	15.33	0.80	0.47	1.33
TL	8.92	3.95	18.99	0.79	0.50	1.20

(1)

ん	わ	ら	や	ま	は	な	た	さ	か	あ	あ段
ひ	び	か	ひ	た	ち	し	き	い	う	う段	
も	ゆ	む	よ	あ	う	つ	く	う	う	う段	
れ	れ	め	へ	へ	て	せ	け	え	え	え段	
ま	う	ま	も	理	の	と	ま	ご	あ	あ段	

(2)

ー	ん	あ	い	う	え	お	が	が	あ段
い	い	い	い	い	い	い	い	い	い段
う	う	う	う	う	う	う	う	う	う段
え	え	え	え	え	え	え	え	え	え段
お	お	お	お	お	お	お	お	お	お段

図6 入力可能な文字

きのキャップを被らせて上向き (NU), 下向き (ND), 右向き (TR), 左向き (TL) の四つ首振り動作をそれぞれ10回ずつ行わせた。そのときのマーカーの動きを2台のビデオカメラで撮影し、マーカーの動きを解析することにより、首振り動作の定量的な値を計測した。縦振りと横振りの動きを観測するために、1台は顔を横側から撮影するように設置し、もう1台は頭上に設置し、マーカーの動きを上側から撮影するように設置した。また、首の動きの動作判定の時間を $T_m = 1.5$ sec, 口の開閉の動作判定の時間を $T_o = 2.0$ sec とした。

マーカーの動きとして、首を振る際の最大回転角度 θ_m および首を振り始めてから元に戻すまでの動作時間 t_m を計測した。計測結果の平均値, 最小値, 最大値を表2に示す。全動作の平均角度は9.57 deg, 平均動作時間は0.78 secであった。

4.3 文入力実験

本実験では、入力可能な文字を図6に示す81字とする文字入力システムを構築した。10人の被験者(A~Iは4.1と同じ人物、Kは成人女性健常者)に対して、下記の5文をそれぞれ入力させたときの、入力時間と入力ミス回数を計測した。

- とっとりだいがく (8字)
- おはようございます (9字)
- おなかがすきました (9字)
- きょうはにちようびです (11字)
- わたしはだいがくせいです (12字)

本実験で使用した計算機はDOS/V PC (CPU:Core2 Duo 2.00GHz)であった。また、口部形状の静止判定の時間は $T_s = 1.0$ sec, 入力候補文字の更新時間は $T_u = 1.0$ sec とした。

文を入力する前に、被験者毎に5種の口部形状をそれぞれ5個ずつ登録した。ただし、4.1の母音形状認識実

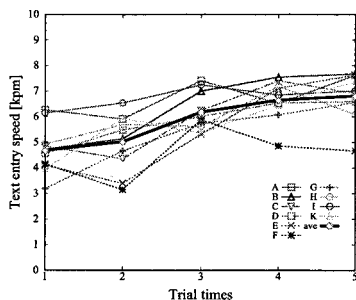


図 7 試行回数と文字入力速度の関係

表 3 文章入力の実験結果

	subjects										ave.
	A	B	C	D	E	F	G	H	I	J	
TES[kpm]	6.74	6.41	5.95	5.83	5.51	4.54	5.24	5.90	6.75	5.81	5.87
MET[%]	7.3	7.7	13.9	3.7	7.0	8.6	9.4	6.9	2.9	13.5	8.1
D^2	0.14	0.00	0.38	0.04	0.07	0.28	0.39	0.58	0.01	0.11	0.20

験では自然な発話を意識させたため、形状の分布が混在しており認識率が低かった。そこで本実験では、母音を意識しつつ、形状の分布が明瞭に分離するようなはっきりとした口部形状を登録させた。この操作は、ユーザが図 2⑥を見ながら形状を作ることにより容易に行える。4.1 で例示した被験者 D と E の特徴量空間の分布を図 5(a) 右、図 5(b) 右に示す。被験者に応じて分布の範囲は異なるが、特に被験者 E は左図と異なり明瞭な分布が観測できる。

入力実験は 5 文を 1 セットとして、5 セットを全て異なる日時で行った。試行回数における入力速度の変化を図 7 に示す。また 5 回の試行における平均入力速度 (text entry speed; TES), 平均入力ミス回数 (miss entry time; MET) を表 3 に示す。1 回目の実験では 1 分間あたりの平均入力文字数 (kana per minute; kpm) は 10 人の平均で 4.7kpm であった。試行回数が増えるたびに文字入力速度があがり、5 回目では 6.8kpm であった。入力速度に個人差があるものの、5 回程度の試行で操作に十分慣れたことがわかる。また最も速い被験者は 5 回目で 7.7kpm であった。また被験者実験のログを解析することにより、文字の入力ミスは 1 回目では 14.1% であったが、5 回目では 4.3% と低下しており、操作に慣れることにより入力ミスがほとんどなくなることを確認した。入力速度および入力ミスが特徴量空間の分布に関係があるか調べるために、登録した学習データに対して誤判別の確率 D^2 [12] を求めた。その結果を表 3 に示すが、誤判別の確率と入力速度、入力ミス回数に相関は見られない。本システムは口部形状を特徴量の推移図を見ながら合わすことができ、その速度は個人差によるものと判断する。概して熟練度によるものと推測する。

5. おわりに

本論文では、口部形状と首の簡単な動きを利用した文

字入力システムを提案した。本手法は、従来のキー操作を利用する方法などに比べ入力速度は速くないが、手指を使わずに人に自然なコミュニケーション手段である発話を意識した新しい文字入力システムである。7~10 名の健常者の協力を得て母音形状認識実験、首振り動作実験および文章入力実験を行い、自然な母音形状では認識精度は 77.4% で低いものの、意識した形状であれば有用であることを確認した。また首振り動作に対しては最大角度差が 10 deg 程度のわずかな動きで十分であることを確認した。約 10 文字の 5 文の文字入力速度は 1 分間に平均 6.8 文字であり、入力ミス率は 4.3% であった。

今後の課題として、ここでは健常者による実験であったが、障害者に対する有用性を評価するための実験を追加する必要がある。また他のインタフェースに比べ文字入力速度が遅い問題があるため、読唇による単語認識技術を取り入れることにより、単語と文字の双方の入力が可能なシステムの提案が挙げられる。

文 献

- [1] 菅原, 新地, 岸野, 小西: “パーソナルコンピュータ上での読唇システムの実時間実現”, 計測自動制御学会論文集, **36**, 12, pp. 1145-1151 (2000).
- [2] K. Iwano, T. Yoshinaga, S. Tamura and S. Furui: “Audio-visual speech recognition using lip information extracted from side-face images”, EURASIP Journal on Audio, Speech, and Music Processing, **2007**, (2007). doi:10.1155/2007/64506.
- [3] 齊藤, 小西: “トラジェクトリ特徴量に基づく単語読唇”, 信学論 (D), **J90-D**, 4, pp. 1105-1114 (2007).
- [4] T. Saitoh, M. Hisagi and R. Konishi: “Analysis of features for efficient japanese vowel recognition”, IE-ICE Trans. Inf. & Syst., **E90-D**, 11, pp. 1889-1891 (2007).
- [5] 織田, 朱, 小沼, 徳野, 未代, 中川: “オフライン識別器を統合したオンライン手書き文字識別器の小型化”, 信学論, **J90-D**, 9, pp. 2583-2594 (2007).
- [6] 尾本, 後藤, 伊藤, 小林: “音声シフト: 音高の意図的な変化を利用した音声入力インタフェース”, 信学論, **J88-D-II**, 3, pp. 469-479 (2005).
- [7] 大野: “視線を用いたインタフェース”, 情報処理論, **44**, 7, pp. 726-732 (2003).
- [8] 伊藤, 数藤, 伊福部: “重度肢体不自由者向けの視線入力方式コミュニケーション装置”, 信学論, **J83-D-I**, 5, pp. 495-503 (2000).
- [9] D. J. Ward and D. J. C. MacKay: “Fast hands-free writing by gaze direction”, Nature, **418**, 6900, p. 838 (2002).
- [10] M. J. Lyons, C.-H. Chan and N. Tetsutani: “Mouthtype: Text entry by hand and mouth”, Proc. of Conference on Human Factors in Computing Systems (CHI2004), pp. 1383-1386 (2004).
- [11] T. Saitoh, M. Hisagi and R. Konishi: “Japanese phone recognition using lip image information”, IAPR Conference on Machine Vision Applications (MVA 2007), pp. 134-137 (2007).
- [12] 奥野: “統多変量解析法”, 日科技連出版社 (1976).