

修士論文

# 運筆音による手書き数字認識



平成 26 年度

三重大学大学院 工学研究科  
博士前期課程 情報工学専攻  
北野 真示



# 目次

<b>第 1 章</b>	<b>序論</b>	<b>5</b>
1.1	はじめに . . . . .	5
1.2	隠れマルコフモデル . . . . .	6
1.3	本稿の構成 . . . . .	6
<b>第 2 章</b>	<b>運筆音の収録</b>	<b>7</b>
2.1	収録内容および条件 . . . . .	7
2.2	運筆音切り出し . . . . .	9
<b>第 3 章</b>	<b>認識実験および結果</b>	<b>13</b>
3.1	状態数についての検討 . . . . .	13
3.2	特徴量についての検討 . . . . .	17
3.3	個人差について . . . . .	30
3.4	認識実験総括 . . . . .	38
<b>第 4 章</b>	<b>結論</b>	<b>41</b>
<b>付録 A</b>	<b>本人データのみを使用した認識について</b>	<b>43</b>
<b>参考文献</b>		<b>45</b>
<b>謝辞</b>		<b>47</b>





# 第 1 章

## 序論

### 1.1 はじめに

人間が生活する社会には様々な音が存在する．音には音声や楽音といったように人間が聞くことを目的とするものがある一方で，人間にとっては雑音とされる音も存在する．しかし，この雑音とされる音をうまく解析することにより，有益な情報を得ることも可能である．近年では，音が収録された場所や状況，収録環境の近くにいるユーザの行動などのシーンを分析する，音響シーン分析技術の研究が行われている．

音響シーンについての先行研究としては，キーボードのタイプ音からの入力テキスト推定 [1] や車両走行時の音からの路面状況 [2] や車両検知 [3]，車種の分類 [4]，音情報からのユーザーに対する危険性の識別 [5]，足音からの歩行者に関する研究 [6][7][8]，実環境音からの行動分類 [9][10][11] などがなされている．

本研究では音響シーンの研究として，運筆音に着目した．運筆音は文字を記述する際に筆記具によって発生する擦過音である．運筆音を対象とした研究としては習字訓練 [12] や運筆音が作業に与える影響 [13]，運筆音からの終筆動作の判別 [14] がなされている．

本研究では運筆音が記述と連動していることから，運筆音から記述した文字の認識は可能ではないかと考えた．文字認識における認識に用いる新たな要素や運筆音から記述の上手さ判定などの応用が考えられる．光学的な文字認識では記述位置とカメラ等の入力機器の間を遮れば認識は不可能になるが，運筆音を遮ることができなければ情報漏えいにつながると予測できる．人間にとっては雑音ととられる運筆音から文字認識を行うことで，情報漏えいの危険性への問題提起へとつながる．この研究で困難な点は，音声認識の研究は多くなされており，どのように状態数や特徴量をとればいいかが確立されているが，運筆音からの認識に関しての研究は存在しないことである．

そこで本研究では，運筆音からの手書き文字認識の最適な手法についての検討を行った．なお，画数が 2 画以内で単純な記述であり，直線だけでなく曲線も適度に含んだ文字として数字を認識の対象にした．被験者に数字 0 から 9 の 10 種類の見本を提示し，この数字の記述で生じる運筆音を研究の対象とした．なお，数字見本の提示については，書き順や文字の大きさの統制をはかるために使用した．認識モデルには隠れマルコフモデル (Hidden Markov Model; HMM) を用いた．

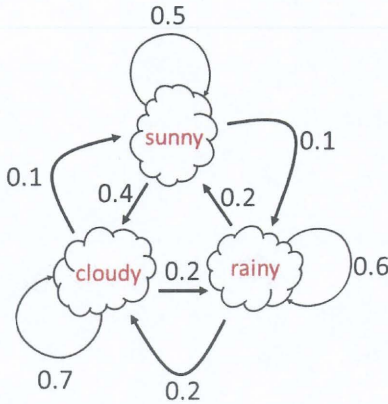


Fig.1.1 マルコフ過程

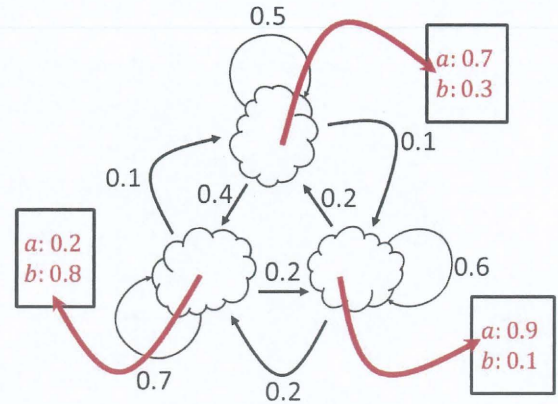


Fig.1.2 隠れマルコフモデル

## 1.2 隠れマルコフモデル

隠れマルコフモデル (Hidden Markov Model; HMM) とは確率モデルの一種であり、音声認識や自然言語処理で用いられる手法である。このモデルの基礎であるマルコフ過程とは次に起こる事象の確率が過去の状態にかかわらず、現在の状態のみで決定する確率過程のことである。Fig.1.1 に示す天気の様子のように過去の天気によらず、現在の天気から未来の天気が確率的に決定するモデルである。また、全ての状態集合  $S$  において、ある状態  $i$  から状態  $j$  に遷移確率は  $p_{ij}$  で表すことができ、次の式を満たす。

$$p_{ij} \geq 0 (i, j \in S); \sum_{j \in S} p_{ij} = 1 (i \in S)$$

隠れマルコフモデルとは、マルコフ過程における状態遷移が隠れている状態であり、観測可能な情報からその状態遷移について推定する手法である。観測できるものが出力信号のみであり、状態遷移については隠れているため、隠れマルコフモデルと呼ばれている。

隠れマルコフモデルの簡単な例を Fig.1.2 に示す。Fig.1.2 の例では Fig.1.1 の天候という状態が隠れ、代わりに確率的に出力される記号  $a, b$  が観測できる。この出力された記号  $a, b$  から状態遷移についての推定を行う。推定を行うアルゴリズムとしては、モデル遷移のパラメータが既知である場合に出力記号から最尤状態列を計算する Viterbi アルゴリズム [15] や出力記号からモデルパラメータを推定する EM アルゴリズム (Expectation Maximization; EM) [16] により構成された Baum-Welch アルゴリズム [17] が存在する。

## 1.3 本稿の構成

本論文の構成としては第2章では数字記述における運筆音の収録方法について述べる。そして、第3章では収録した運筆音を用いた認識方法の検討とその結果について述べる。最後に第4章では本稿の結論を述べる。

## 第2章

# 運筆音の収録

### 2.1 収録内容および条件

運筆音が文字の記述に依存するため、自由に記述を行うとデータにばらつきが生じると考えられる。運筆音からの手書き数字認識を考える際にこのばらつきを低減する必要がある。文字の大きさや書き順、記述にかかる時間を統制する目的で、Fig.2.1 に示す数字の練習プリント [18] を参考に用意した数字見本を用いて行った。なお、被験者に提示した見本は 2.5 cm 四方である。被験者にはこの数字見本を意識しつつ自然な記述を行うことを依頼した。記述の手順としては数字 0 から 9 の順でそれぞれ 10 回の記述を 3 回行った。収録した音はこの記述により生じた音であり、収録後に一字ごとに切り出しを行った。

収録は被験者 24 名（右利き 23 名，左利き 1 名）で，3 つの時期に分けて行った。1 回目の収録は，4 名を対象に，大学実験室 (5.38 × 4.31 × 3.50 m) で実施した。この際の暗騒音レベルは 31.7 dB(A) であった。2 回目と 3 回目の収録はそれぞれ，被験者 5 名（右利き 4 名，左利き 1 名）と 15 名で大学会議室 (7.00 × 7.93 × 2.70 m) で行った。この際の暗騒音レベルはそれぞれ 37.4 dB(A) と 34.8 dB(A) であった。実際の収録風景に関しては Fig.2.2 に示す。なお，Fig.2.2 は 3 度目の収録風景である。

3 回の収録に使用した機器とその接続に関して Fig.2.3 に示す。収録は PC(Aspire one D150Bk73, acer)，USB オーディオ・インタフェース (UA-1G USB Audio Interface, Roland)，マイクアンプ (AT-MA2, audio-technica)，無指向性マイクマイク (ECM-77B/9x, SONY) を順に接続して行った。マイクの設置位置は机の上方 10 cm である。なお，記述に使用した文具は，鉛筆 (三菱鉛筆事務用鉛筆 9800, HB) とコピー用紙 (SDG70A4R, シスコムネクスト) である。この収録機器構成において，サンプリング周波数 44.1 kHz，量子化ビット数 16 bit で運筆音を収録した。なお，振幅の正規化については行っていない。

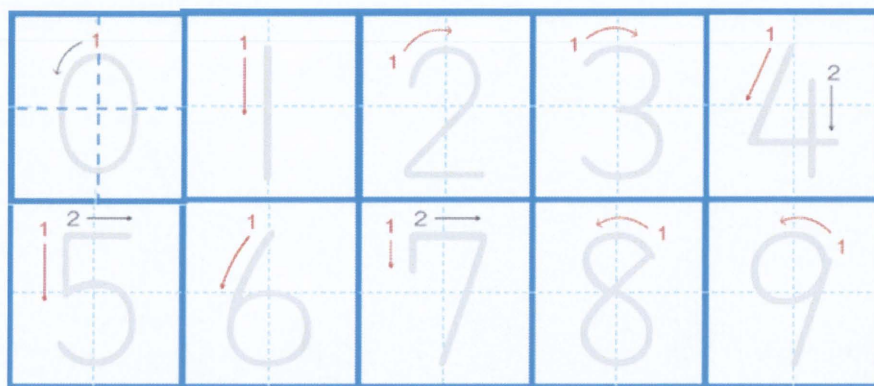


Fig.2.1 数字見本

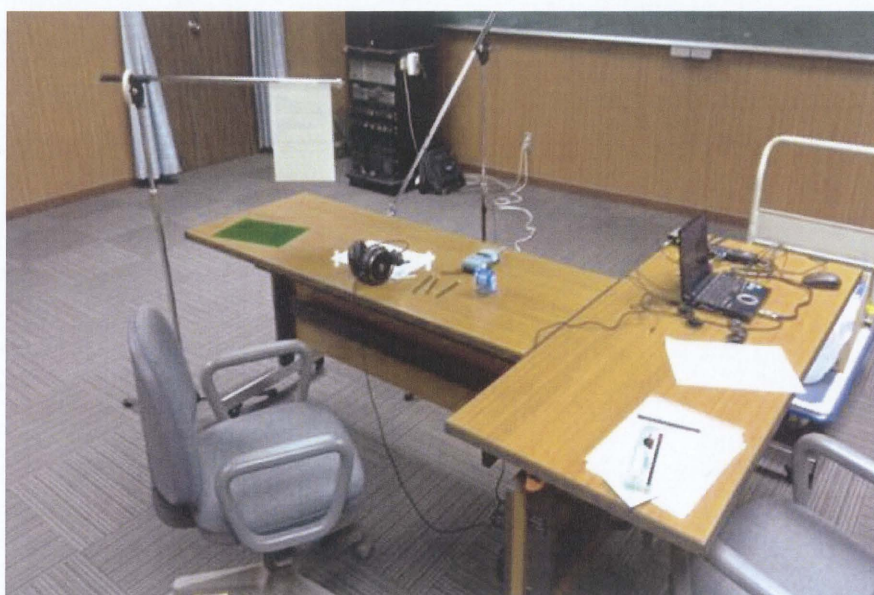


Fig.2.2 収録風景

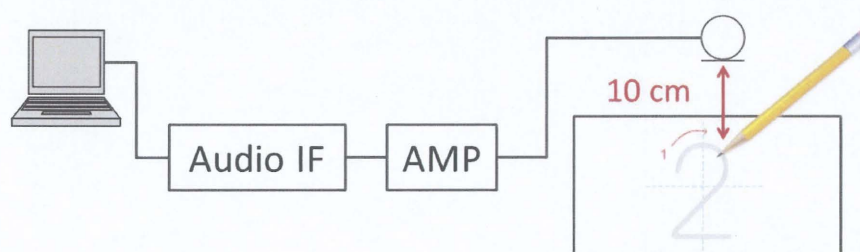


Fig.2.3 収録機器の配置図

## 2.2 運筆音切り出し

収録された運筆音データには、鉛筆の接触音、記述音、次の数字へ移る際の移動音が含まれる。Fig.2.4 の上部に記述時の動作、下部にこの波形の運筆音の波形とそのスペクトログラムを示す。なお、スペクトログラムの濃淡はパワーの強さを示す。スペクトログラムにおいて、実線が記述音に相当し、その下枠の数字は記述と対応している。破線で示した部分が接触音に相当し、二重線で示した部分が次の数字へ移る際に発生した音に相当する。また、数字7の二画目からもわかるように、一画目から二画目に移るときだけでなく数字中の角でも記述の一時停止が判別可能であることがわかった。

また、数字記述において、『撥ね』や『払い』が表れることもあった。Fig.2.5 に撥ねと払いの一例を示す。Fig.2.5 は撥ねと払いについて実際の記述およびそのスペクトログラムである。スペクトログラム中の黒枠が撥ねと払いに相当する。このような撥ねや払いが一画の終わりに現れることがあった。

本研究では、これらの音のうち、接触音と記述音を運筆音として用いた。また、撥ねおよび払いも運筆音として含めた。接触音を含めたのは、数字7のような2画の数字では2画目の初めに接触音が含まれることから、接触音を除外することが困難であること、また記述時の特徴量として利用できないかと考えたためである。Fig.2.4 の場合は図中の赤線部分での切り出しとなる。

このような規則に基づき、収録した運筆音データから一文字ごとの波形データを切り出し、認識実験に用いた。波形の切り出しは、フリーソフトウェアの WavePad [20] を用い、手作業にて行った。なお、切り出しを行った数字0から9のスペクトログラムの一例を Fig.2.6 に示す。スペクトログラム下の枠内の数字は記述に相当する。被験者1名当たりの運筆音は、各数字30回のうち不備データを除いた各数字25回、被験者1名当たり合計250の運筆音を認識実験で使用した。



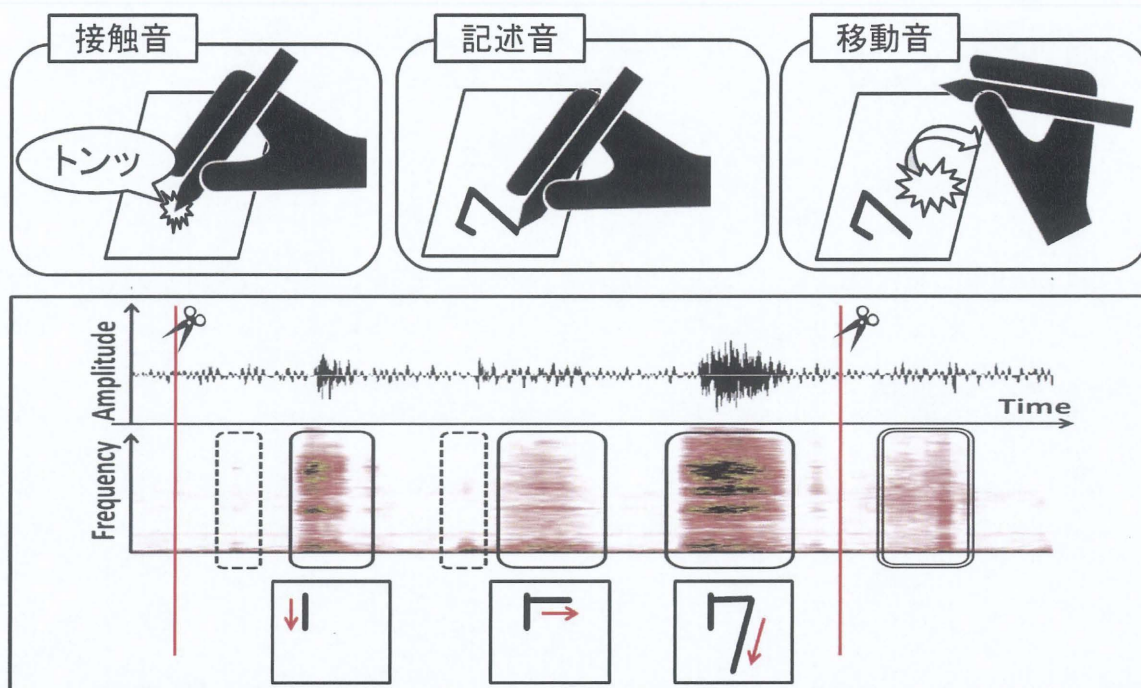


Fig.2.4 運筆音の切り出し 記述時の動作（上），収録音の一例（下）

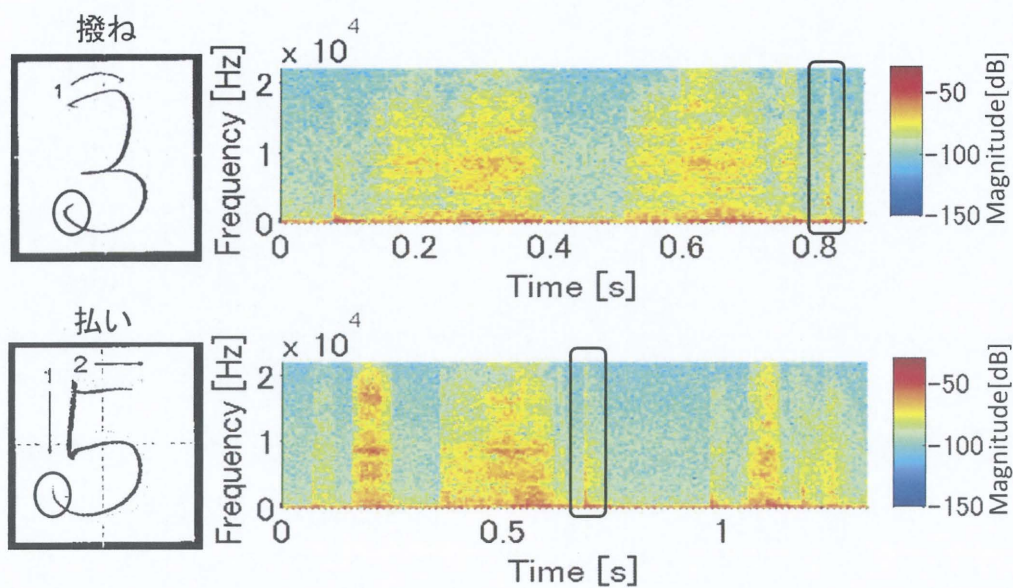


Fig.2.5 撥ねと払い

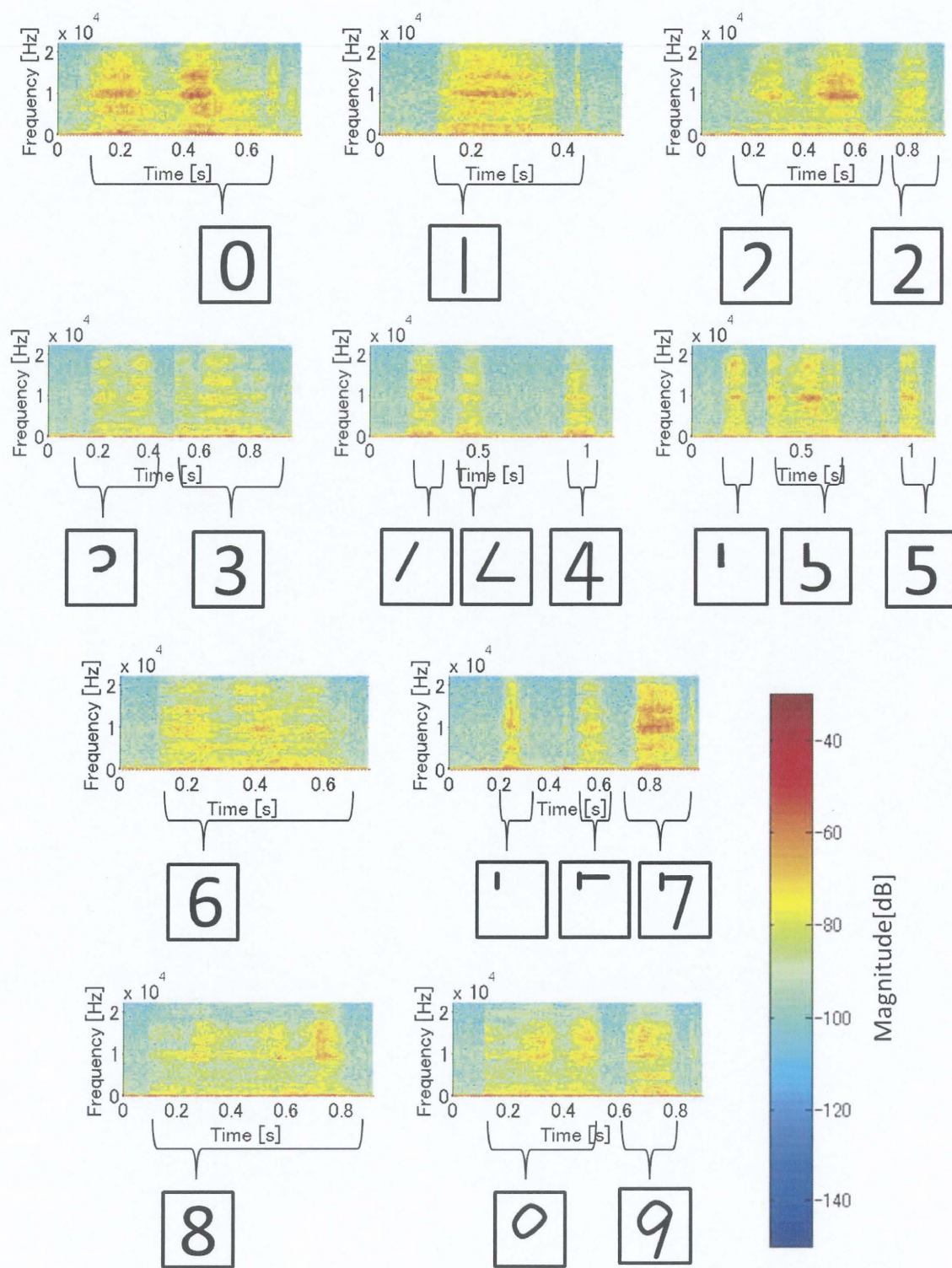


Fig.2.6 運筆音の一例





## 第3章

# 認識実験および結果

運筆音から手書き数字を認識するために、本研究では音声認識などの分野で用いられている隠れマルコフモデル (Hidden Markov Model; HMM) による認識器を用いた。なお、実験では HMM のツールキットである HTK Ver.3.4.1[19] を用いた。

認識実験では右利き被験者 23 名のうち 10 名分の運筆音を用いて行った。この運筆音データを用いて運筆音からの手書き数字認識について調査を行った。なお、運筆音データは一回目の収録分が 2 名、二回目の収録分が 3 名、三回目収録分が 5 名である。

学習には 1~9 名分の被験者のデータをそれぞれ用い、学習に用いなかった被験者のデータを認識対象として、実験を行った。なお、実験は被験者データ全ての組合せに対して行った。幾つかの観点から学習認識を行い、運筆音からの手書き数字認識の最適な方法について検討した。

### 3.1 状態数についての検討

#### 3.1.1 認識条件

HMM による音声認識においては、国際音声記号など、規格化された音素があり、認識したい単語ごとに音素列を記述することが可能である。しかし、運筆音についてはこのような音素記号や、その状態数についての定義は無い。

そのため、認識実験を行うに当たり数字ごとの状態数についての検討を行った。右利き被験者 23 名の平均記述時間を Fig.3.1 に示す。この平均記述時間より、各数字ごとに状態数を振り分けた。状態数の振り分けとしては記述時間 0.5 s ごとに 2 から 5 状態を振り分け、認識実験を行った。

なお、認識に用いる特徴量については音声認識で使用する特徴量を参考にした。Table.3.1 に使用した特徴量を示す。メル周波数ケプストラム係数 (Mel-Frequency Cepstrum Coefficient; MFCC)、微分、加速度を用いた。用いた次数については 0 から 12 次の 13 次をそれぞれ用いた 39 次である。

#### 3.1.2 認識結果および考察

認識実験の結果を Fig.3.2, Fig.3.3 および Fig.3.4 に示す。Fig.3.2 は学習データ数の推移による平均認識率、Fig.3.3 と Fig.3.4 はそれぞれ学習データ数 8 名の際の被験者ごとの認識率と数字ごとの認識率である。なお、凡例の数字は認識に用いた 0.5 s あたりの状態数である。

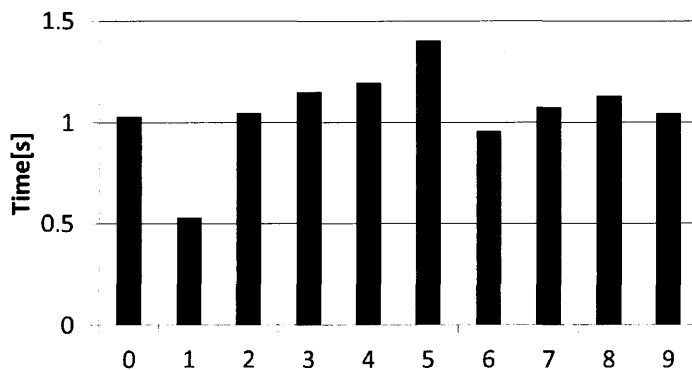


Fig.3.1 数字ごとの平均記述時間（各数字 575 回の平均）

Table.3.1 使用した特徴量について

特徴量	MFCC + $\Delta$ MFCC + $\Delta\Delta$ MFCC
次数	それぞれ 0-12 次
メルフィルタバンク数	26
リフタリング値	22
フレームの幅	25 ms
フレームシフト長	10 ms
窓関数	ハミング窓

Fig.3.2 より、用いる状態数を増やすことで認識率は向上した。また、学習データ数を増やすことで認識率は向上した。しかし、状態数が 0.5 s あたり 4 状態の場合と 5 状態の場合とでは差があまり生じなかった。学習データ数ごとに見た 4 状態と 5 状態の認識率の差は 0.5 % 以下であった。この結果より、平均記述時間から状態数を割り当てる場合、少なくとも 0.5 s あたり 4 状態が適切と考えた。

また、Fig.3.3 と Fig.3.4 より被験者ごとや数字ごとに認識率のばらつきがあることがわかった。特に 0.5 s あたり 4 状態の認識で被験者 B と H、数字 0 と 6 で 80 % 以下の認識率となった。0.5 s あたり 4 状態、学習データ数 8 名の認識の内訳については Table.3.5 に示す。Table.3.5 より、数字 0 と 6 では互いに誤認識される場合が多くある。これは数字 0 と 6 の曲線の急峻以外の書き方が似ているため、運筆音からの認識が困難であると考えられる。なお、被験者ごとの認識率の差については後の 3.3 節で述べる。

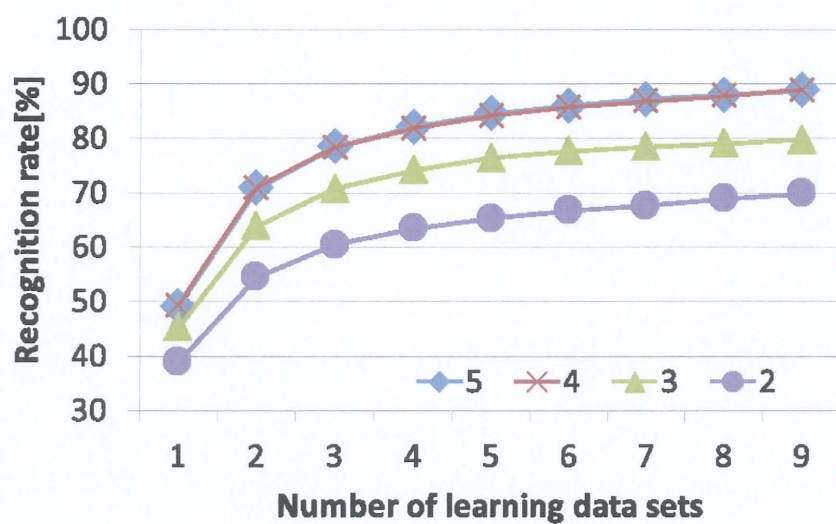


Fig.3.2 各状態数による学習データ数ごとの平均認識率

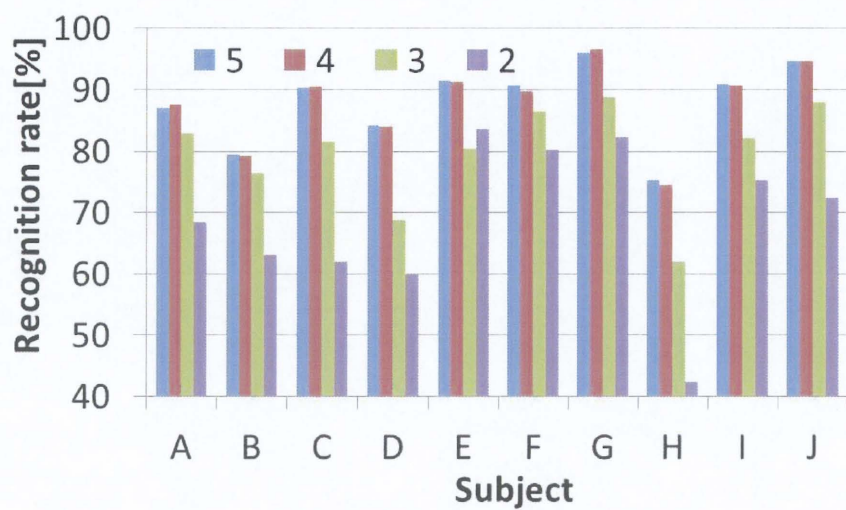


Fig.3.3 各状態数による被験者ごとの平均認識率 (学習データ数 8 名)

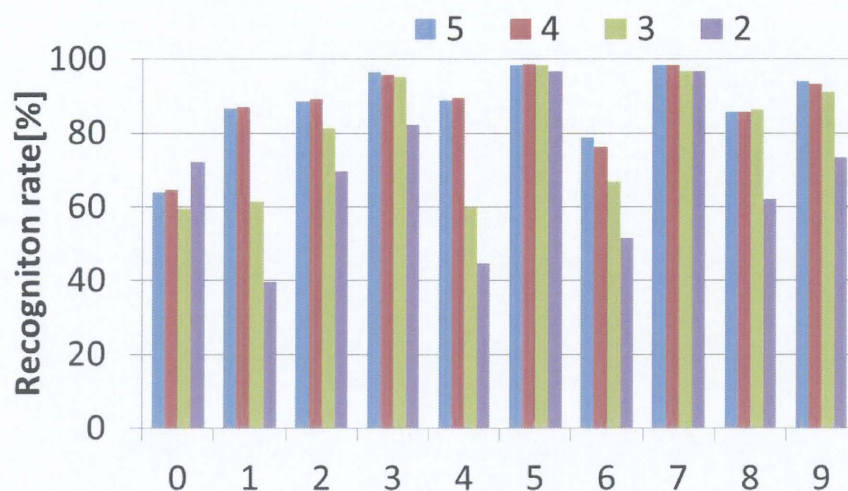


Fig.3.4 各状態数による数字ごとの平均認識率 (学習データ数 8 名)

Table.3.2 0.5 s ごとに 4 状態を用いた場合の認識結果の内訳

Target digits	Recognition rate [%]									
	0	1	2	3	4	5	6	7	8	9
0	64.76	6.98	1.20	1.60	0	0	23.51	0.76	1.20	0
1	2.85	86.98	2.67	0	2.49	0.09	2.18	2.40	0	0.36
2	1.24	0.09	89.24	1.51	2.13	0.18	0.80	0.76	0.49	3.56
3	0.53	0.67	1.64	95.96	0.04	0	0.49	0.40	0.18	0.09
4	0.49	0.36	0.40	0	89.47	7.20	0.13	1.42	0	0.53
5	0.04	0	0	0	0.44	98.80	0.71	0	0	0
6	11.16	8.13	1.11	0.22	0.18	0.27	76.18	0.44	1.87	0.44
7	0.44	0	0	0	1.29	0	0	98.27	0	0
8	4.45	0.44	0.40	3.20	0	0	3.60	0.40	85.73	1.78
9	0.36	0	0.71	0	0	0	0.13	0.22	5.16	93.42

## 3.2 特徴量についての検討

運筆音からの手書き数字認識を行うに当たり、適切な特徴長についての検討を行った。検討の対象は以下の4パターンである。

1. 特徴量の次数変更
2. 時間変動要素の有無
3. 0次成分の有無
4. 特徴量の種類の変更

なお、状態数については数字ごとの平均記述時間 Fig.3.1 より、0.5 s ごとに4状態を持つと仮定して行った。

### 3.2.1 特徴量の次数変更

特徴量に用いる次数を変更した際の認識率について調査を行うに当たり、使用した特徴量を Table.3.3 に示す。Table.3.1 より、次数部分のみを変更し、0 から6次の場合と0 から12次の場合、0 から18次の場合で認識実験を行った。なお、特徴量については MFCC およびその微分・加速度を用いた。

結果を Fig.3.5, Fig.3.6 および Fig.3.7 に示す。Fig.3.5 は学習データ数の推移による平均認識率、Fig.3.6 と Fig.3.7 はそれぞれ学習データ数8名の際の被験者ごとの認識率と数字ごとの認識率である。凡例の0-6 は特徴量の0 から6次を使用した結果であり、同様に0-12, 0-18 は特徴量の0 から12次, 0 から18次を使用した結果である。

結果として、学習データ数ごとの結果には大きな差が生じていないが、数字ごとの結果に関しては用いる次数によって差が生じていた。数字0 と4 では使用する特徴量の次数が増加するとともに認識率が向上した。そして、数字2 と8 では使用する特徴量の次数が増加するとともに認識率が低下した。

結論として、特徴量の次数を変更することで数字ごとの正答に変化が生じる。しかし、全体の認識率では認識率に大きな差は生じないことを確認した。以降の認識実験では0 から12次を基本として認識実験を行う。

Table.3.3 使用した特徴量について

特徴量	MFCC + $\Delta$ MFCC + $\Delta\Delta$ MFCC
次数	それぞれ 0-6 次 それぞれ 0-12 次 それぞれ 0-18 次
メルフィルタバンク数	26
リフタリング値	22
フレームの幅	25 ms
フレームシフト長	10 ms
窓関数	ハミング窓

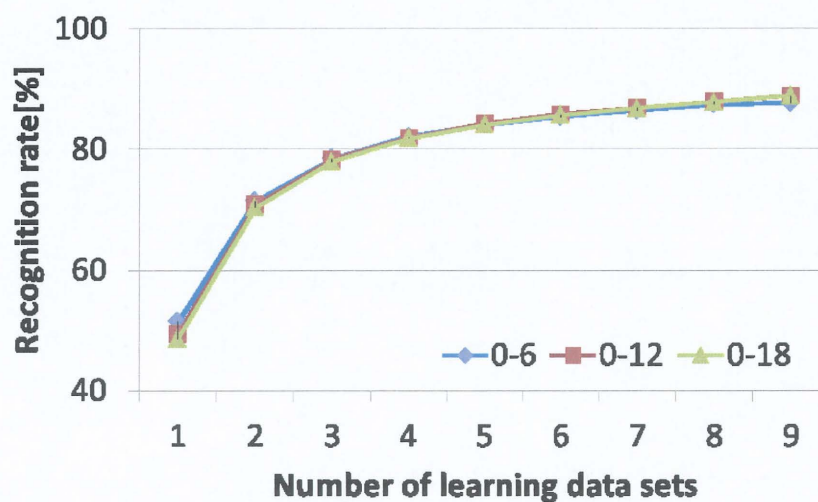


Fig.3.5 学習データ数ごとの平均認識率



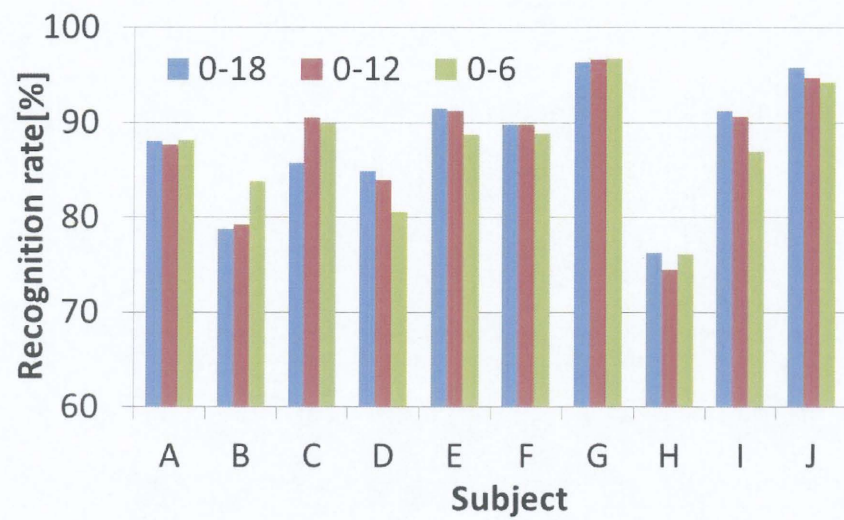


Fig.3.6 被験者ごとの平均認識率 (学習データ数 8 名)

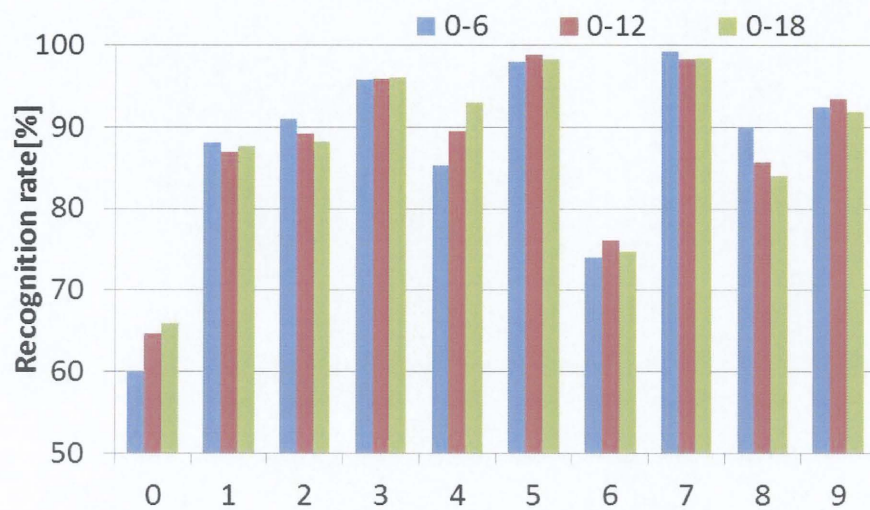


Fig.3.7 数字ごとの平均認識率 (学習データ数 8 名)

### 3.2.2 時間変動要素の有無

時間変動要素である微分・加速度の有無による認識率についての調査を行うに当たり、使用した特徴量を Table.3.4 に示す。対象とする時間変動要素である微分・加速度の有無で3条件で認識実験を行った。条件は微分と加速度を含む場合、微分のみを含む場合、時間変動要素を含まない場合である。なお、特徴量は MFCC で0から12次を使用した。

結果を Fig.3.8, Fig.3.9 および Fig.3.10 に示す。Fig.3.8 は学習データ数の推移による平均認識率、Fig.3.9 と Fig.3.10 はそれぞれ学習データ数8名の際の被験者ごとの認識率と数字ごとの認識率である。凡例の MFCC\_0 は MFCC の0から12次の場合、MFCC\_0D は MFCC の0から12次と微分の場合、MFCC\_0DA は MFCC の0から12次と微分・加速度の場合の認識率を示している。

Fig.3.8 より、全ての学習データ数で時間変動要素である微分・加速度を含む認識の方が MFCC のみの認識より正答を得ることができた。時間変動要素を加えることで約10～20%の認識率向上を確認した。ただし、微分の場合と微分と加速度の場合とでは、認識率に大きな差は生じてはいなかった。両者の認識率の差は最大でも5%未満であった。

Fig.3.9 より、被験者ごとの認識において MFCC\_0DA が MFCC\_0D に比べて被験者 D と H 以外で高い結果となった。また、被験者によって時間変動の考慮による認識率の向上具合が異なることを確認した。被験者 C と H, J では20%以上の向上が見られた一方で、被験者 D では10%以下の向上となった。

Fig.3.10 より、数字ごとの認識においては数字1以外で時間変動要素を加えた場合の方が認識率が向上する結果となった。時間変動要素を含まない場合の認識の内訳についてを Table.3.5 に示す。Table.3.5 より、他の数字が数字1に誤認識される場合が多く存在することがわかる。特に数字0, 6, 8 で20%以上の数字1への誤認識が生じた。

この数字0, 6, 8 についてはいくつかの共通点が存在する。まず、一画の数字であり、数字中に角もなく記述中に一時停止が存在しないこと、そして数字が曲線で構成されていることである。この共通点より、時間変動要素は画や角などの記述の一区切りでその運筆音がどのような線を描いたかをより良く認識させる。つまり、記述の一字停止の数が同じ文字での誤認識を抑制する効果があると考えられる。

結論として、時間変動要素である微分・加速度は運筆音からの文字認識を行う上で有効であることを確認した。特に、曲線を持つ数字の認識精度向上に寄与した。



Table.3.4 使用した特徴量について

特徴量	MFCC + $\Delta$ MFCC + $\Delta \Delta$ MFCC MFCC + $\Delta$ MFCC MFCC
次数	それぞれ 0-12 次
メルフィルタバンク数	26
リフタリング値	22
フレームの幅	25 ms
フレームシフト長	10 ms
窓関数	ハミング窓

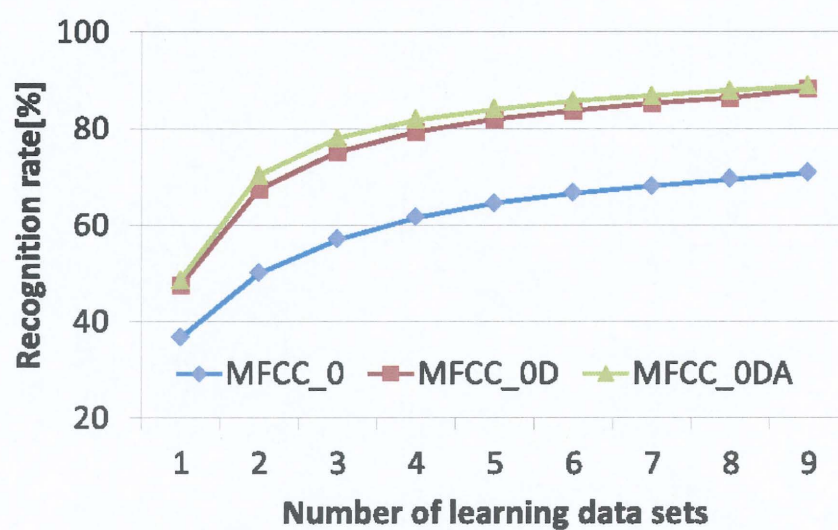


Fig.3.8 学習データ数ごとの平均認識率

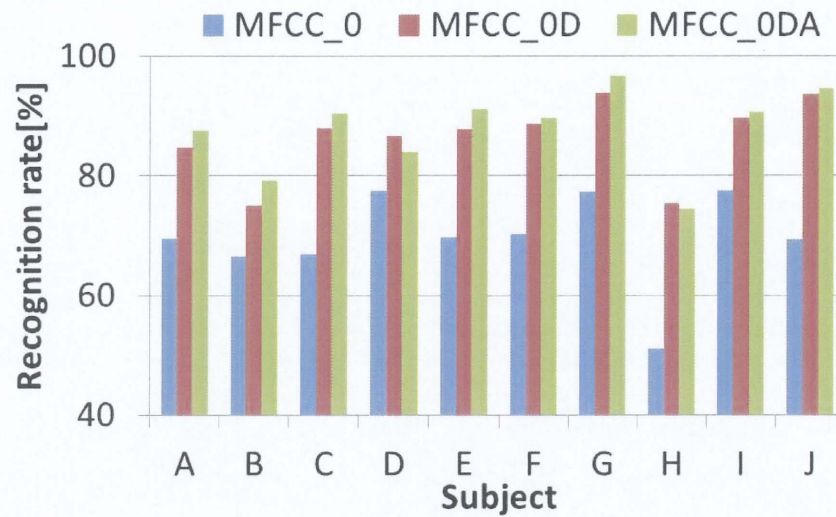


Fig.3.9 被験者ごとの平均認識率 (学習データ数 8 名)

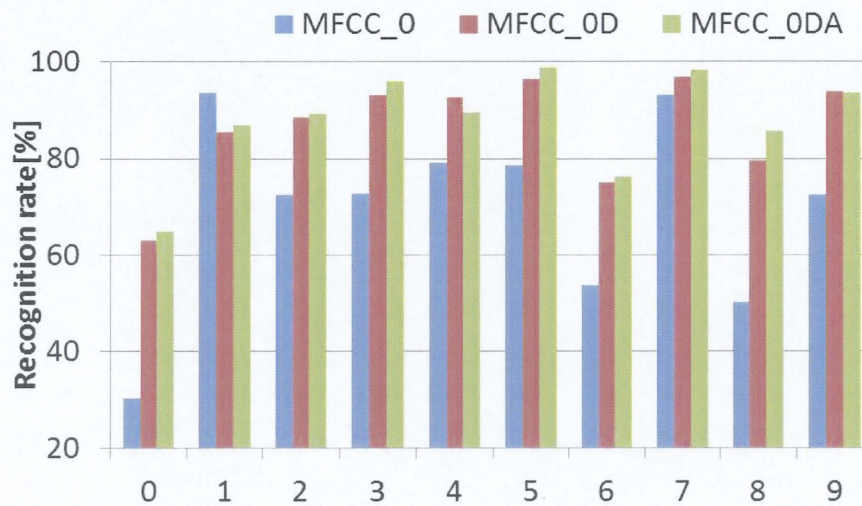


Fig.3.10 数字ごとの平均認識率 (学習データ数 8 名)

Table.3.5 時間変動要素なしの認識結果の内訳

Target digits	Recognition rate [%]									
	0	1	2	3	4	5	6	7	8	9
0	30.31	30.80	1.07	11.38	0.40	0.13	24.62	0	0.71	0.58
1	0.18	93.51	1.42	0	3.02	0.13	0.49	0.53	0	0.71
2	0.49	7.95	72.45	5.42	2.80	1.16	1.78	1.07	0.40	6.49
3	3.02	6.53	9.24	72.71	0.44	0.22	3.69	0.76	1.33	2.04
4	0.04	2.67	0.84	0.13	79.07	10.45	0.04	6.58	0	0.18
5	0	1.38	3.24	0.13	15.11	78.58	0.09	1.02	0.04	0.40
6	12.58	25.16	2.22	0.98	0.93	0.18	53.69	0.22	1.91	2.13
7	0.22	2.00	0.09	0.09	4.04	0.27	0.18	92.98	0	0.13
8	7.56	24.35	2.58	6.31	0.09	0	5.78	0.13	50.13	3.07
9	1.47	7.56	8.93	1.64	1.73	0.09	0.18	1.38	4.67	72.36

### 3.2.3 0 次成分の有無

運筆音のパワーに相当する 0 次成分の有無による認識率の調査を行うに当たり，使用した特徴量を Table.3.6 に示す．0 次成分を含む 1 から 12 次の場合と含まない 1 から 13 次の場合の 2 条件認識実験を行った．合計次数は 2 条件ともに微分・加速度を含めて 39 次である．

結果を Fig.3.11, Fig.3.12 および Fig.3.13 に示す．Fig.3.11 は学習データ数の推移による平均認識率，Fig.3.12 と Fig.3.13 はそれぞれ学習データ数 8 名の際の被験者ごとの認識率と数字ごとの認識率である．凡例の 0-12 は 0 から 12 次を使用した結果，1-13 は 1 から 13 次を用いた認識結果である．

Fig.3.11 より，全ての学習データ数で 0 次成分を含む 0 から 12 次の結果が 1 から 13 次と比較して高い認識結果となった．学習データ数ごとに見ると 3 から 5 % ほどの上昇が確認できた．また，被験者ごと数字ごとの認識結果 Fig.3.12, Fig.3.13 では，被験者 B のみで認識率が低下したが，その他の被験者と数字で 0 次成分を含む場合の認識率向上が確認できた．

結論として，全ての数字で 0 次成分を含むことで平均認識率は上昇を示したことより，0 次成分が運筆音の認識において必要な要素と考えた．

Table.3.6 使用した特徴量について

特徴量	MFCC + $\Delta$ MFCC + $\Delta \Delta$ MFCC
次数	それぞれ 0-12 次 それぞれ 1-13 次
メルフィルタバンク数	26
リフタリング値	22
フレームの幅	25 ms
フレームシフト長	10 ms
窓関数	ハミング窓

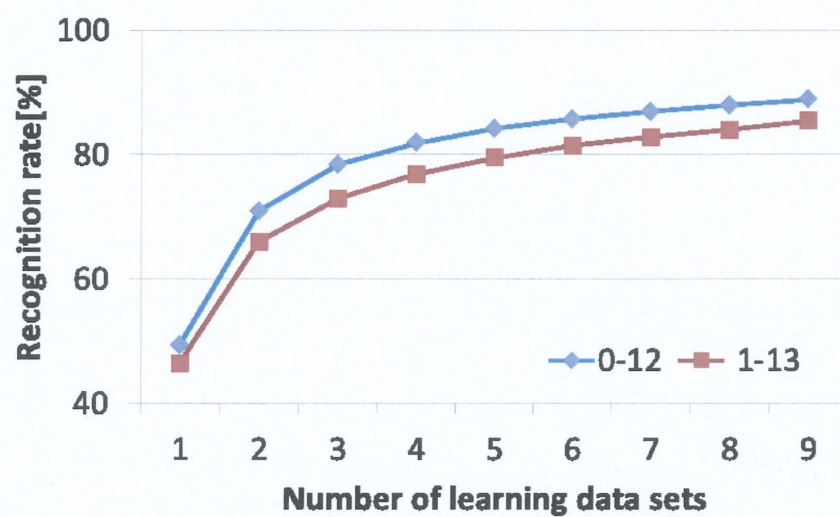


Fig.3.11 学習データ数ごとの平均認識率

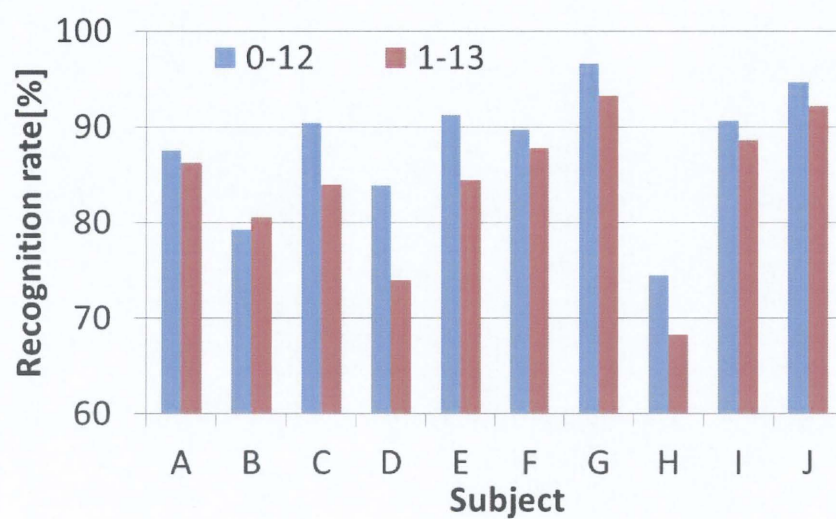


Fig.3.12 被験者ごとの平均認識率 (学習データ数 8 名)



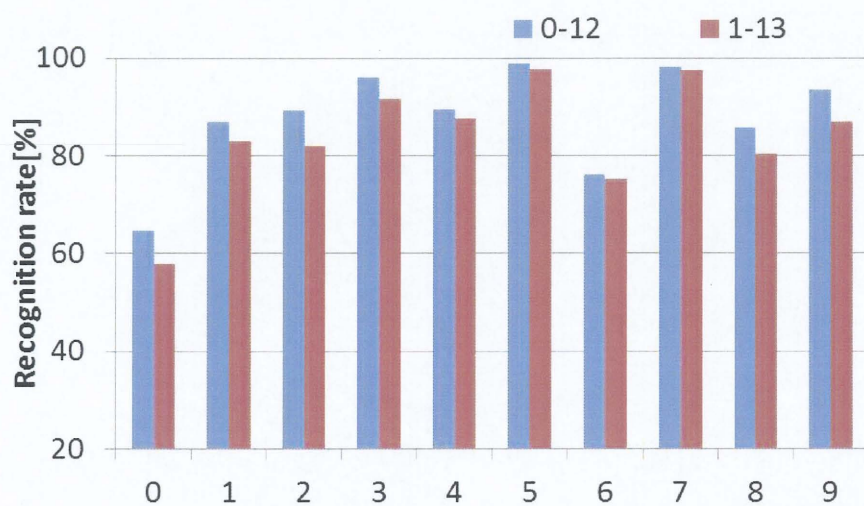


Fig.3.13 数字ごとの平均認識率 (学習データ数 8 名)

Table.3.7 使用した特徴量について

特徴量	MFCC + $\Delta$ MFCC + $\Delta \Delta$ MFCC LPC + $\Delta$ LPC + $\Delta \Delta$ LPC LPCep + $\Delta$ LPCep + $\Delta \Delta$ LPCep
次数	0-12 次, E + 1-12 次
メルフィルタバンク数	26
リフタリング値	22
フレームの幅	25 ms
フレームシフト長	10 ms
窓関数	ハミング窓

### 3.2.4 特徴量の種類の変更

MFCC を対象として認識実験を行ってきたが、これが適切であるか検討するために MFCC 以外の特徴量を用いた認識実験を行った。用いた特徴量は線形予測フィルタ係数 (Linear prediction filter coefficients; LPC), LPC ケプストラム係数 (LPC cepstrum coefficient; LPCep) である。特徴量についての詳細を Table.3.7 に示す。ケプストラムである MFCC と LPCep は 0 から 12 次, LPC はエネルギー成分 (=E) と 1-12 次を次数として用いた。全ての条件で合計次数は 39 字である。

認識結果を Fig.3.14, Fig.3.15 および Fig.3.16 に示す。Fig.3.14 は学習データ数の推移による平均認識率, Fig.3.15 と Fig.3.16 はそれぞれ学習データ数 8 名の際の被験者ごとの認識率と数字ごとの認識率である。

Fig.3.14 より, 学習データ数 2 名以上の認識で MFCC, LPCep, LPC の順に高い認識率となった。学習データ 9 名の際の平均認識率は MFCC で 88.9 %, LPC で 79.4 %, LPCep で 84.1 % であった。

Fig.3.15 および Fig.3.16 より, 被験者・数字ごとの平均認識率でも Fig.3.14 の結果と同様に MFCC, LPCep, LPC の順に高い認識となるものが大半であった。しかし, 一部の被験者・数字ではその限りではなく, 特に被験者 D, F および H や数字 0 では MFCC 以外の特徴量を用いた結果が高い認識となった。特に, この被験者ごと際に関しては記述による被験者ごと個人差と考える。

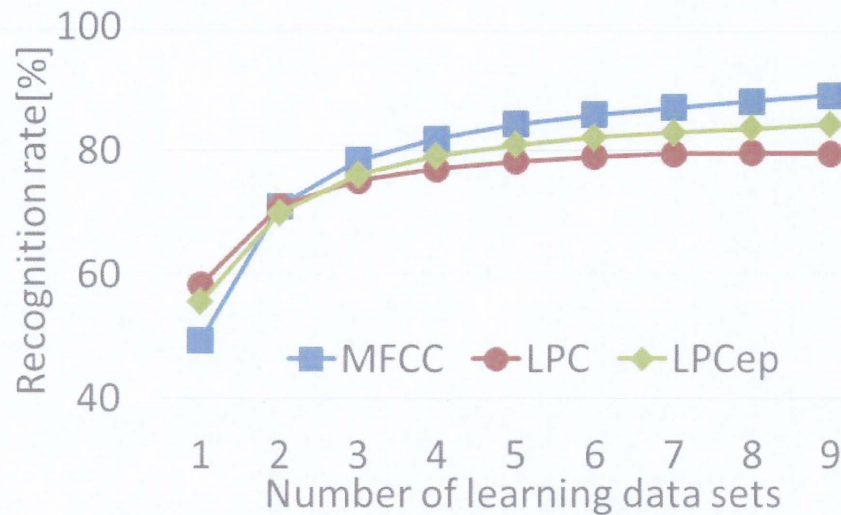


Fig.3.14 学習データ数ごとの平均認識率

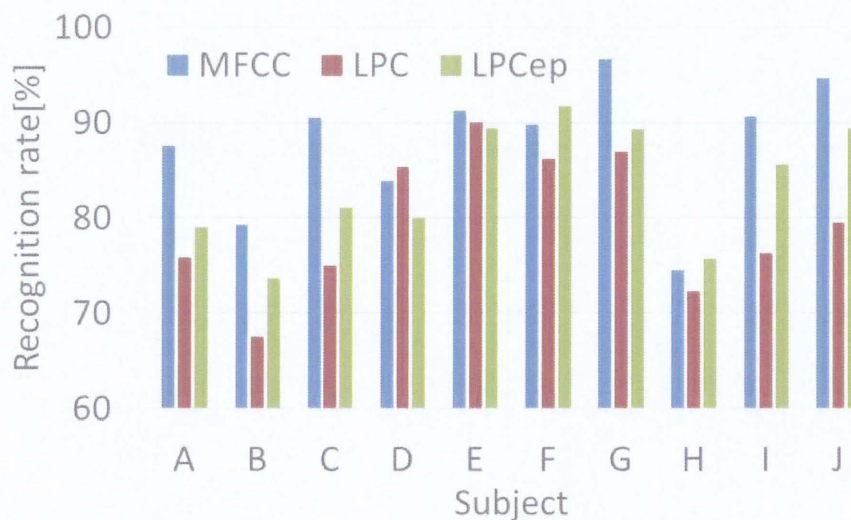


Fig.3.15 被験者ごとの平均認識率 (学習データ数 8 名)

### 3.2.5 特徴量について結果と考察

運筆音からの手書き数字認識の最適な特徴量の検討を行った。認識実験の結果として、MFCC、LPC および LPCep の 3 種類の特徴量で認識実験を行った結果、MFCC を用いた平均認識率が最も高い結果となった。また、運筆音のパワーに対応する 0 次成分や時間変動要素である微分・加速度が運筆音による認識で有効であることを確認した。しかし、次数を変更した際には平均認識率に大きな差が確認できなかった。

以上より結論として、特徴量に MFCC とその微分・加速度を用いて、0 次成分を含むことが運筆音からの認識で有効である。



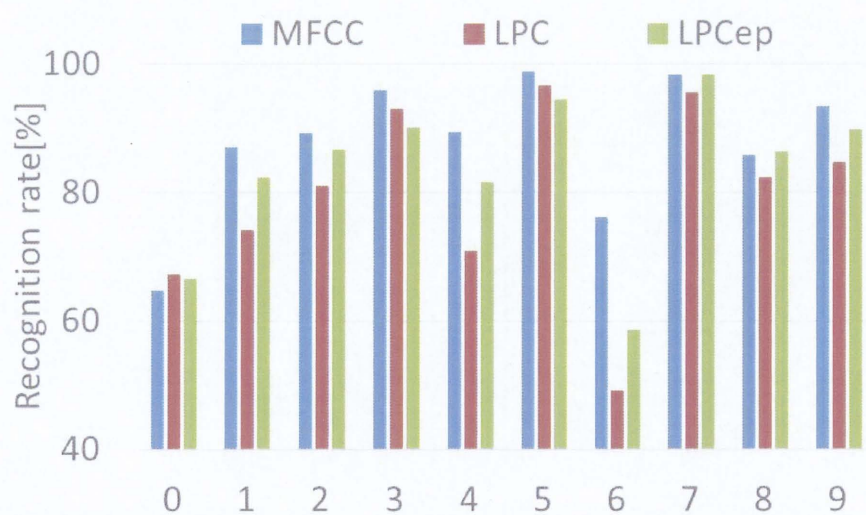


Fig.3.16 数字ごとの平均認識率 (学習データ数 8 名)

### 3.3 個人差について

3.1 節および 3.2 節で行った認識実験において被験者ごとの認識率に差が生じていた。この認識率の個人差についての調査を行った。これまでに使用した 10 名の被験者において特に被験者 H が低い認識率を示した。ここから被験者 H に着目し、被験者 H と他の被験者の差について調査した。

被験者 H と被験者全体の認識結果について考えた。Fig.3.17 に被験者 H と被験者全体の数字ごとの認識率、Table.3.8 に被験者 H の認識の内訳を示す。なお、Fig.3.17 および Table.3.8 の認識は Table.3.1 の条件で状態数は 0.5 s ごとに 4 状態で行った認識の学習データ数 8 名の際の結果である。Fig.3.17 より、被験者 H と被験者全体の認識結果を比較すると特に数字 6 で低い認識率となった。また、認識の内訳である Table.3.8 より、被験者 H の認識は数字 0 と 6 が 50 % 以上数字 1 へと誤認識された。この誤認識の傾向は 3.2.2 節の時間変動要素の有無における結果と似ている。ここから、被験者 H に発生した認識率の差は時間変動が上手く学習されていないために起こったのではないかと考えた。

これを考えるに当たり、被験者 H と被験者全体の記述時間についての調査を行った。Fig.3.18 に被験者 H と被験者全体の平均記述時間を示す。Fig.3.18 より、被験者 H の記述は被験者全体と比較して、全ての数字記述で速いことがわかった。なお、被験者ごとの平均記述時間を Table.3.9 に、数字 0 から 9 の被験者ごとの記述時間のヒストグラムを Fig.3.19 から Fig.3.28 に示す。数字ごとのヒストグラムより、被験者ごとに各数字の記述時間はある程度収束していることがわかる。また、被験者 H の次に低い認識率となっていた被験者 B の平均記述時間を見ると被験者 H とは逆に記述が遅いことがわかった。

結論として、筆圧や癖字など他の要因も考えられるが、記述速度は認識率の個人差に影響していると考えた。この記述速度による個人差を抑制するためにはより多くのサンプルの使用などが考えられる。また、今回の実験では学習と認識で使用する被験者データを分けて行った。このためある被験者を認識データとして用いる場合、その被験者データは学習データとして使用することはなかった。認識に用いる被験者データの一部を認識データとして用いれば認識率の向上につながる可能性もある。本人データのみを使用した認識を付録 A で述べる。

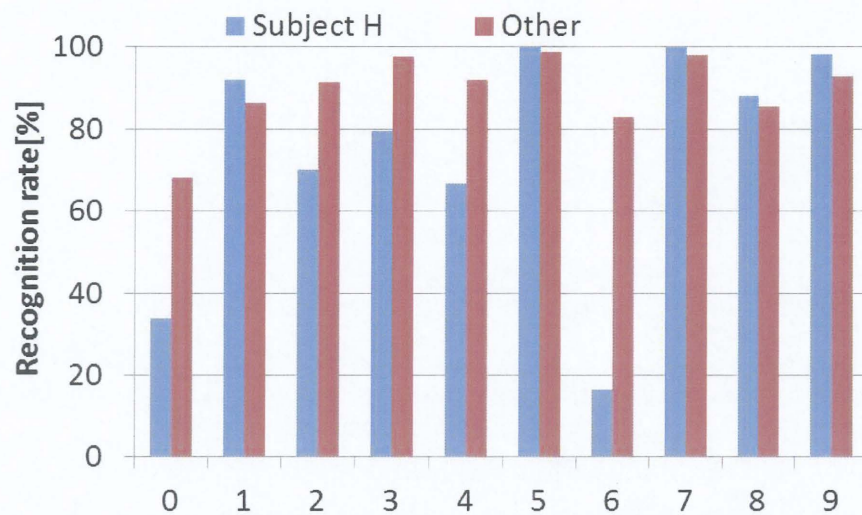


Fig.3.17 被験者 H とそれ以外の被験者の数字ごとの認識率 (学習データ数 8 名)

Table.3.8 被験者 H の認識の内訳

Target digits	Recognition rate [%]									
	0	1	2	3	4	5	6	7	8	9
0	33.78	52.44	0	0	0	0	1.78	7.56	4.44	0
1	0	92.00	0	0	4.00	0	0	4.00	0	0
2	0	0	70.22	0	21.33	0.44	0	0.89	0	7.11
3	0	0.89	14.22	79.56	0.44	0	0	4.00	0	0.89
4	0	0	0	0	66.67	33.33	0	0	0	0
5	0	0	0	0	0	100	0	0	0	0
6	5.78	68.89	3.11	0	1.33	0	16.44	4.00	0	0.44
7	0	0	0	0	0	0	0	100	0	0
8	0.89	2.22	3.11	0	0	0	0	4.00	88.00	1.78
9	0	0	0	0	0	0	0	1.78	0	98.22

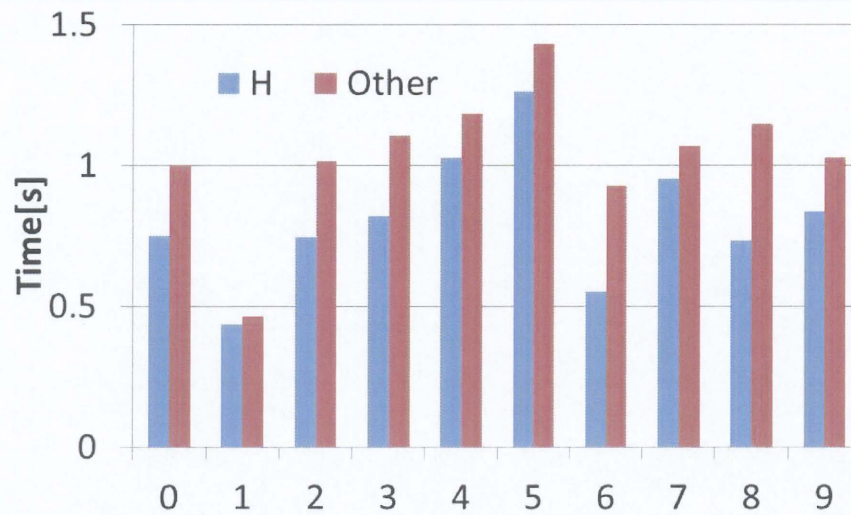


Fig.3.18 被験者 H とそれ以外の被験者の平均記述時間

Table.3.9 被験者の平均記述時間

	Digits									
	0	1	2	3	4	5	6	7	8	9
A	1.11	0.45	1.12	1.21	1.37	1.45	0.96	1.05	1.19	1.12
B	1.65	0.74	1.46	1.68	1.56	1.91	1.62	1.66	1.88	1.57
C	1.31	0.42	1.23	1.14	1.11	1.41	0.94	1.13	1.33	1.06
D	0.82	0.43	1.06	1.15	1.21	1.64	0.89	0.96	1.06	0.97
E	0.81	0.27	0.73	0.87	1	1.26	0.72	0.96	0.96	0.84
F	0.75	0.36	0.74	0.8	1.05	1.17	0.67	0.96	0.93	0.84
G	0.68	0.41	0.88	1.06	1.05	1.34	0.78	0.91	0.94	0.9
H	0.75	0.44	0.74	0.82	1.02	1.26	0.55	0.95	0.73	0.84
I	1.05	0.56	1.02	1.07	1.2	1.45	0.99	1.07	1.18	1.04
J	0.78	0.54	0.88	0.95	1.12	1.24	0.77	0.92	0.85	0.89
Average	0.97	0.46	0.99	1.08	1.17	1.41	0.89	1.06	1.11	1.01



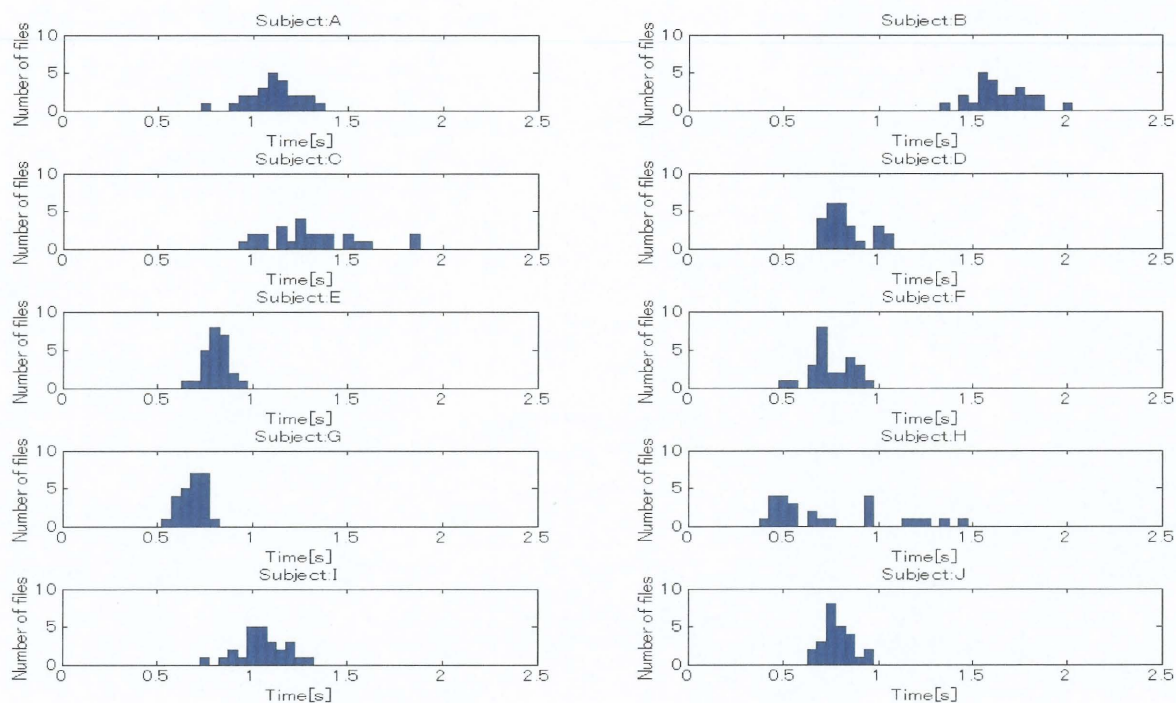


Fig.3.19 数字0 の記述時間のヒストグラム

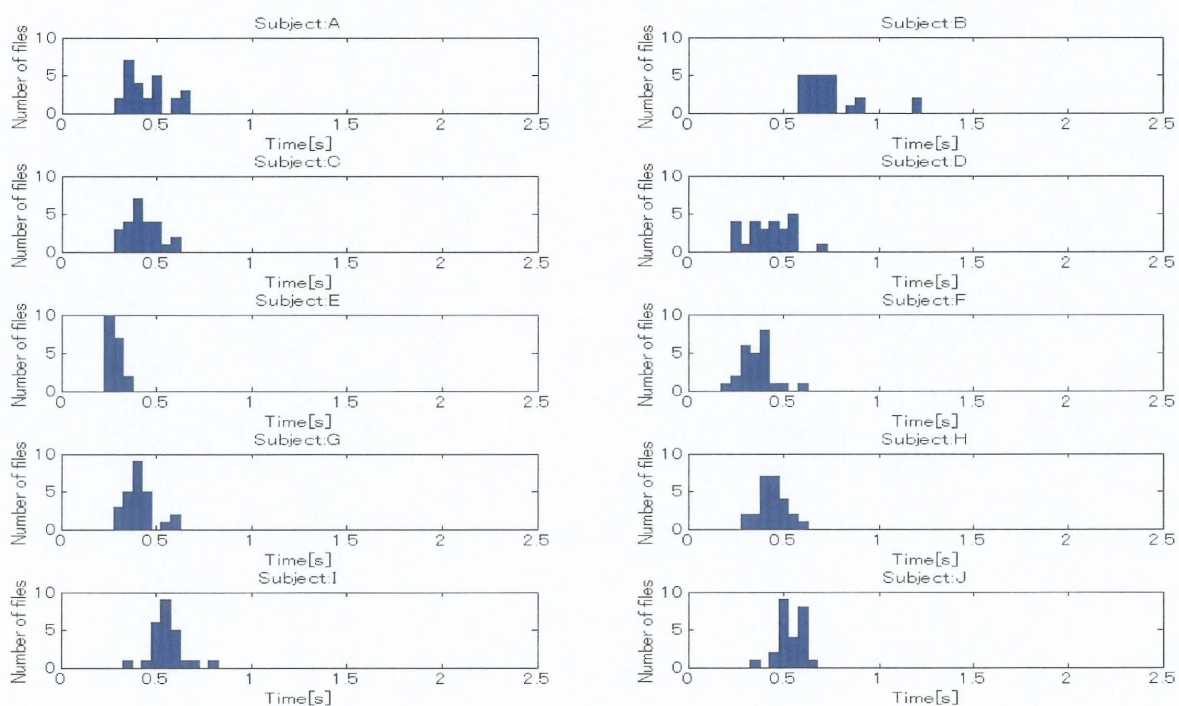


Fig.3.20 数字1 の記述時間のヒストグラム

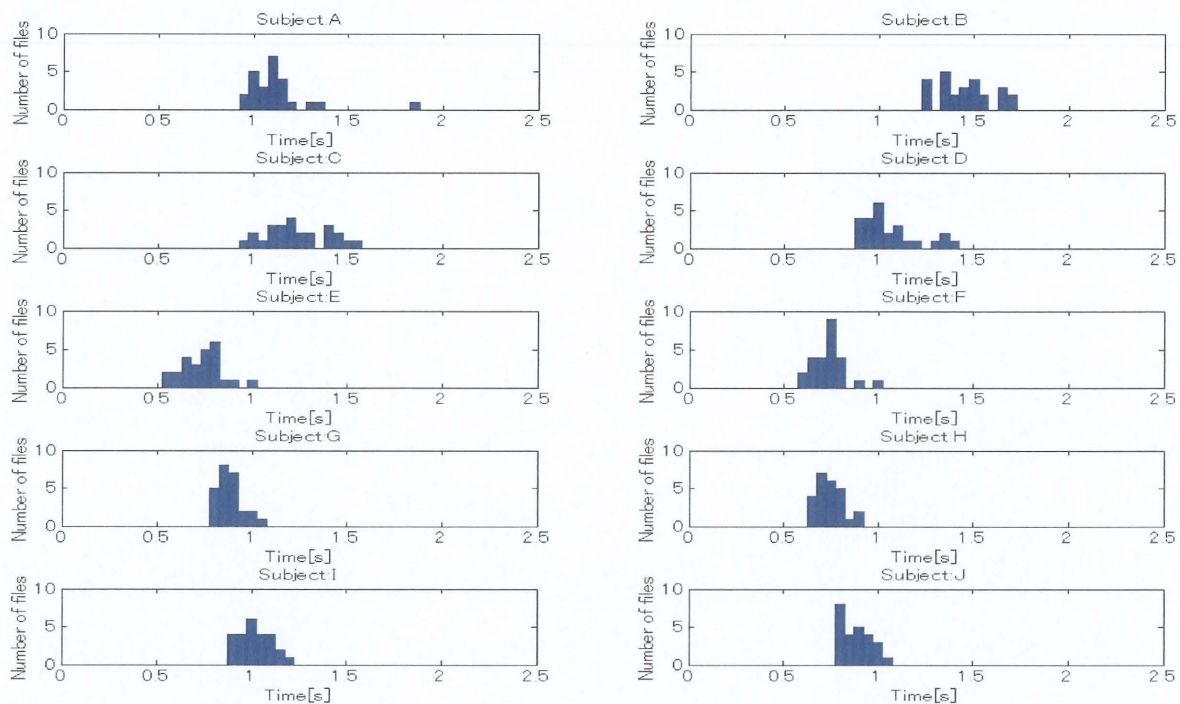


Fig.3.21 数字 2 の記述時間のヒストグラム

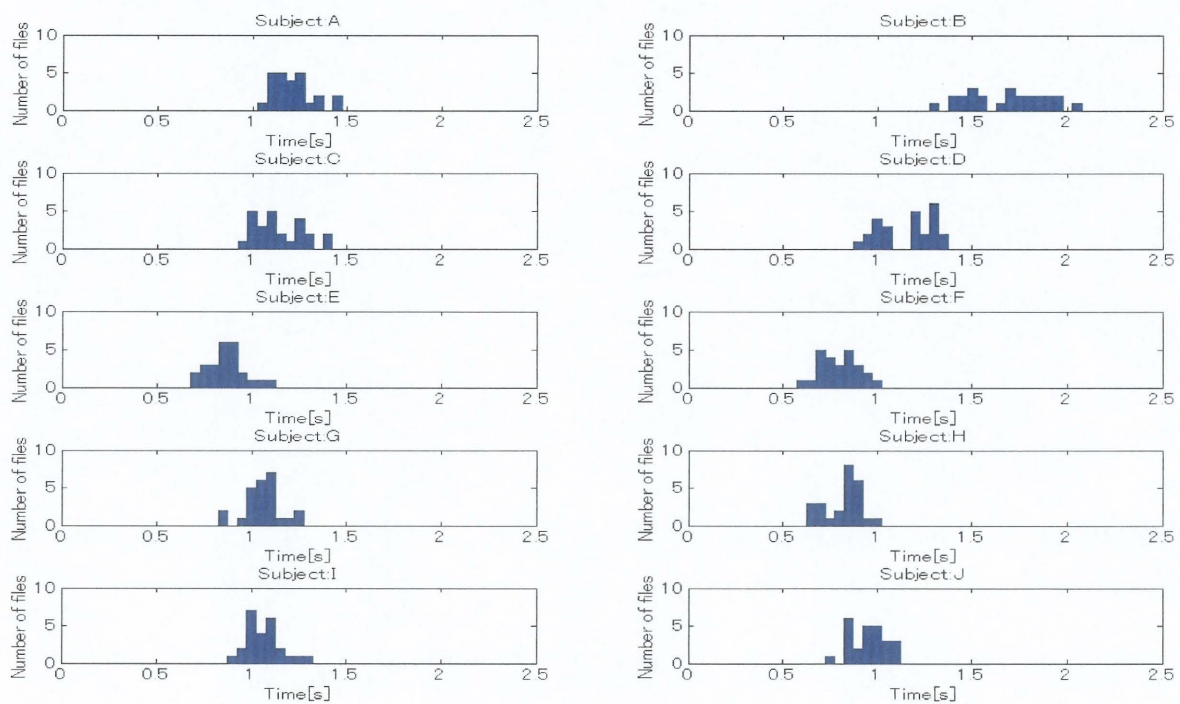


Fig.3.22 数字 3 の記述時間のヒストグラム

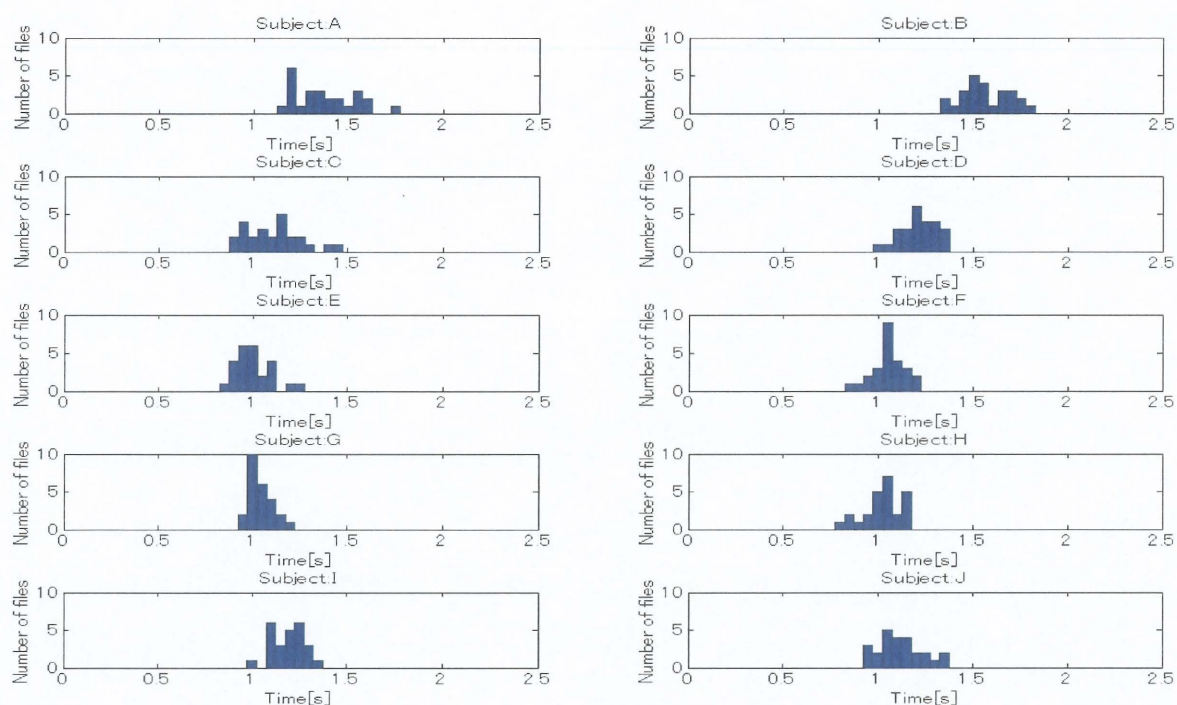


Fig.3.23 数字 4 の記述時間のヒストグラム

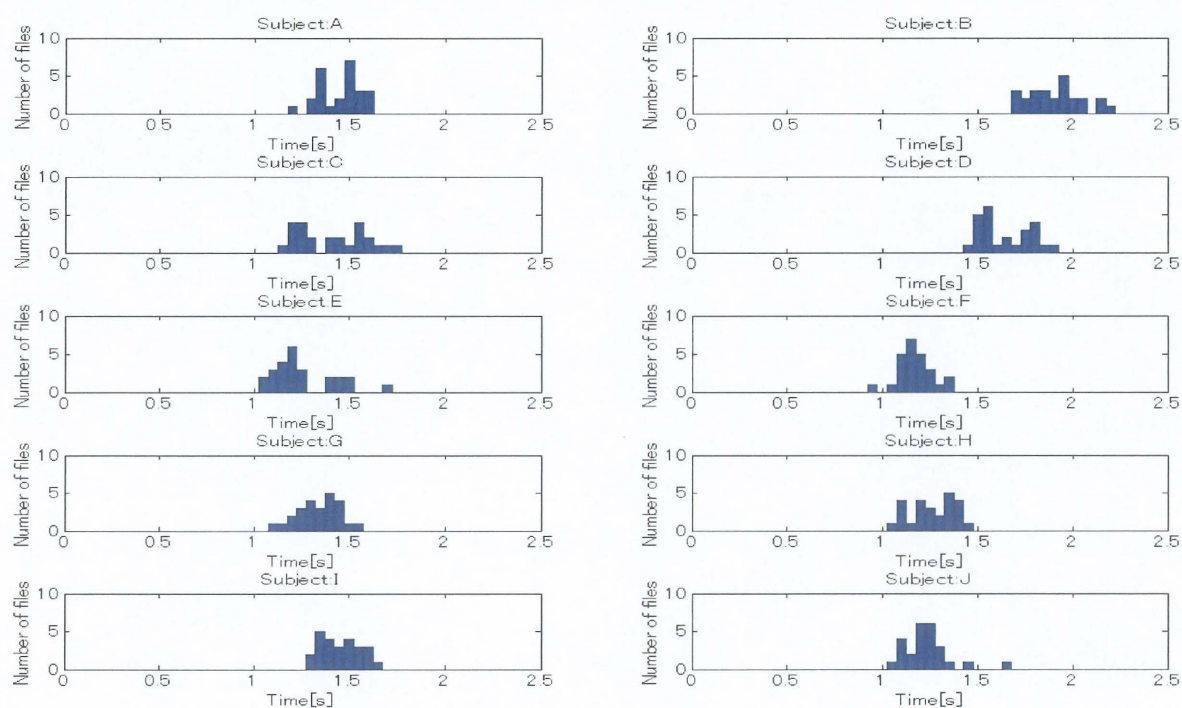


Fig.3.24 数字 5 の記述時間のヒストグラム



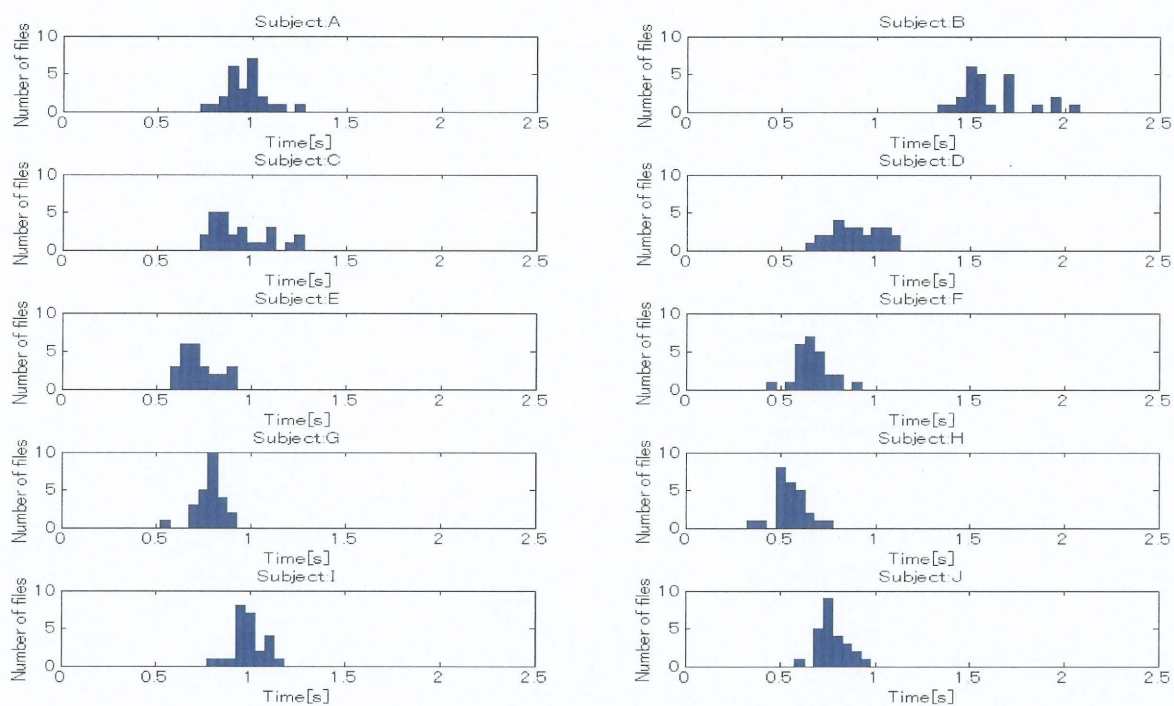


Fig.3.25 数字6 の記述時間のヒストグラム

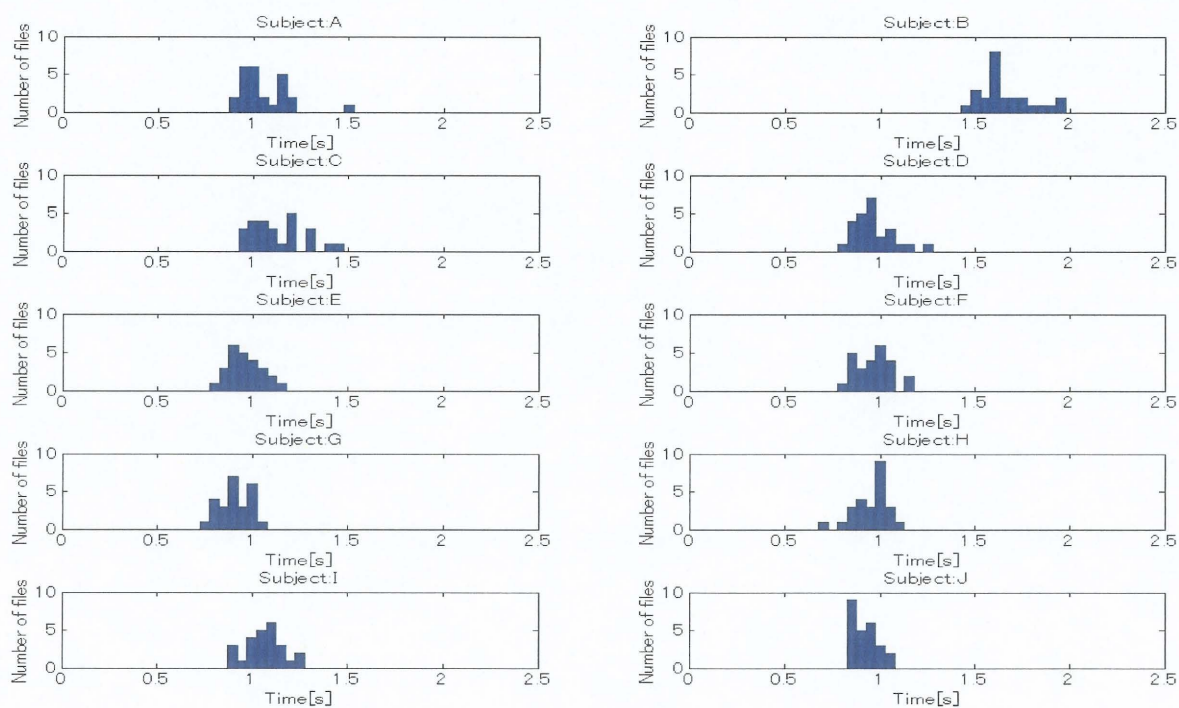


Fig.3.26 数字7 の記述時間のヒストグラム



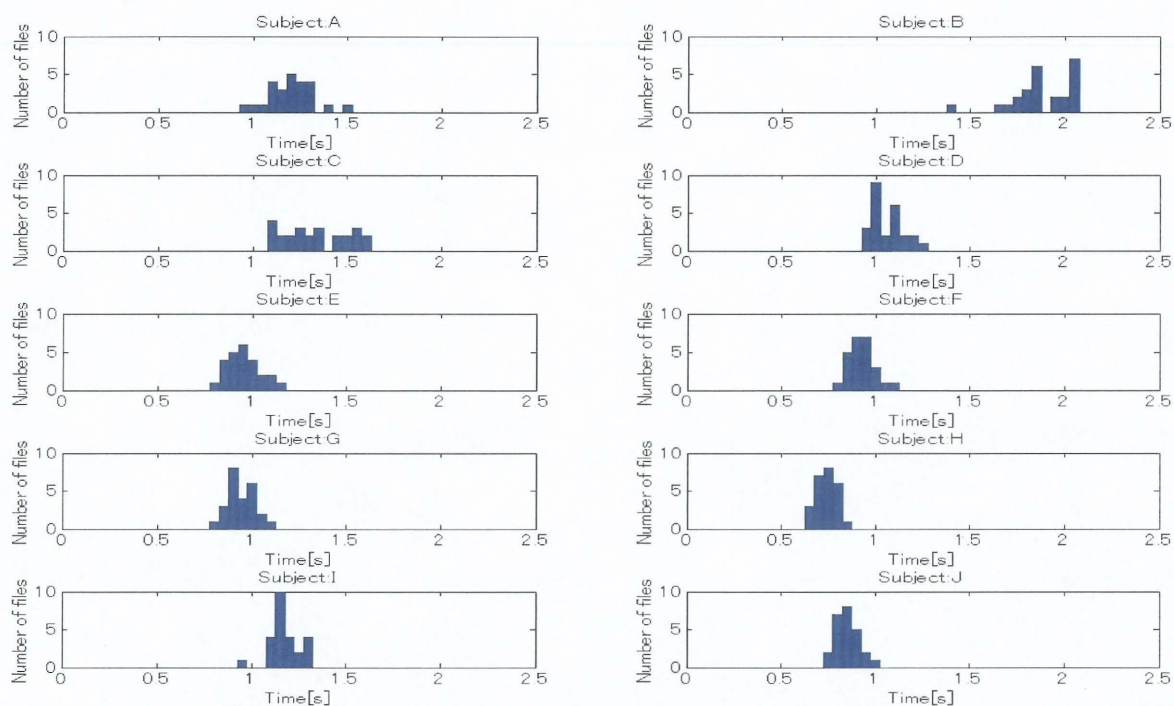


Fig.3.27 数字8の記述時間のヒストグラム

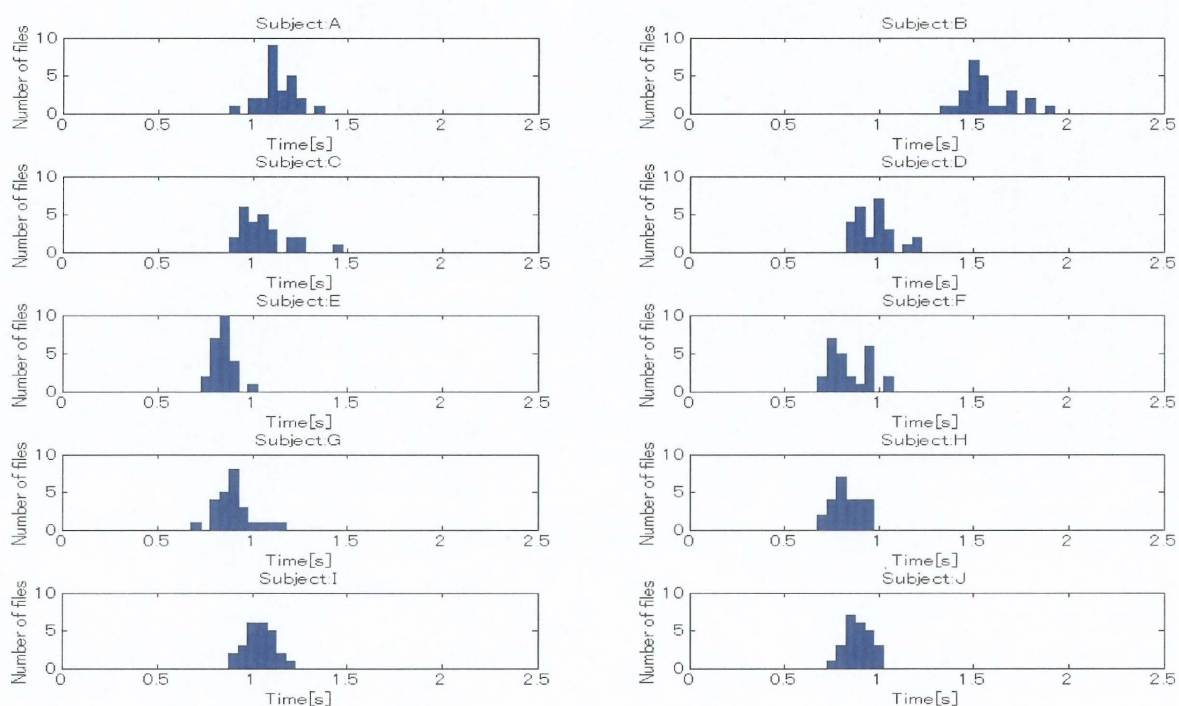


Fig.3.28 数字9の記述時間のヒストグラム

### 3.4 認識実験総括

被験者 10 名分の運筆音データを用いて、最適な HMM の状態数、特徴量についての検討を行った。この結果で確認した最適な特徴量を用いた収録した右利き被験者全員、23 名による認識実験を行った。HMM の状態数は平均記述時間より 0.5 s ごとに 4 状態、特徴量に関しては Table.3.1 と同様である。認識方法に関しては 10 名による認識と同様に、学習データとして使用しなかった被験者データを認識データとして用いた。なお、学習・認識の組合せは膨大な数になるため、学習データ数 1 から 3 名および 19 から 22 名の際の平均認識率を算出した。

認識実験の結果を Fig.3.29, Fig.3.30 および Fig.3.31 に示す。Fig.3.29 は学習データ数ごとの平均認識率である。凡例の Min, Max および Average は、各学習データ数における認識率の最小値、最大値および平均値を示す。Fig.3.30 は学習データ数 21 名における被験者ごとの平均認識率である。Fig.3.31 は 23 名による実験で学習データ数 21 名と 10 名による実験で学習データ 9 名の数字ごとの平均認識率である。凡例はそれぞれ用いた学習データ数を示す。

Fig.3.29 より、平均認識率は学習データ数 19 から 22 名で約 88 % となった。これは 10 名による認識の学習データ 9 名と同様の数値であった。この結果より、本稿で検討した条件において、平均認識率はおよそ 9 割の正答が上限であると考ええる。

Fig.3.30 より、学習に用いるデータ数を増やしても被験者ごとの平均認識率にばらつきが生じることを確認した。この結果より、記述速度等のバリエーション増加や学習データに本人データ用いるなど個人差の対策が必要であると考ええる。また、被験者 23 名と 10 名の認識実験における数字ごとの平均認識率を示す Fig.3.31 より、被験者 23 名の方が数字 0, 2, 3, 5, 6, 7, 8 で認識率の上昇を確認した。特に数字 0, 6, 8 の認識率上昇より、被験者数を増やすことで曲線の認識精度向上が示唆されたと考える。

結論として、本稿で検討した運筆音による手書き数字認識の上限はおよそ 9 割の正答であると考えた。この結果以上の正答を得るためには、運筆音のバリエーションや認識条件の変更が必要であると考ええる。

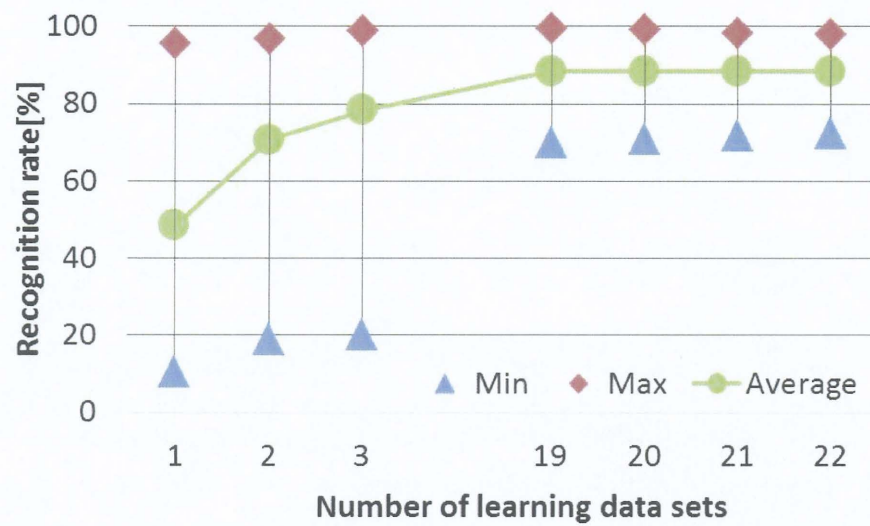


Fig.3.29 学習データ数ごとの平均認識率

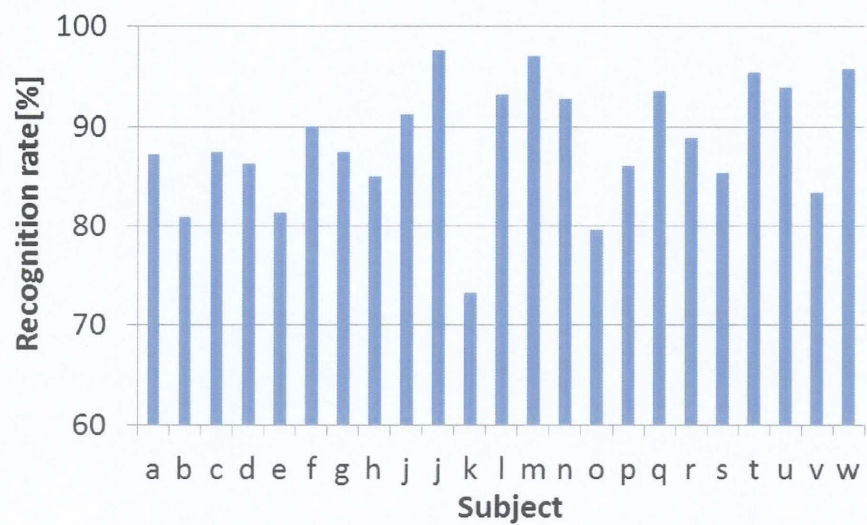


Fig.3.30 被験者ごとの平均認識率 (学習データ数 21 名)

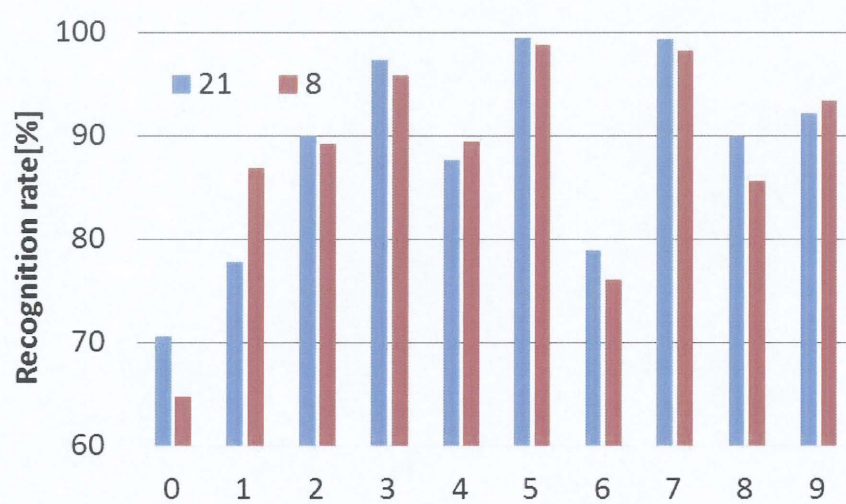


Fig.3.31 数字ごとの平均認識率 (被験者 23 /学習 21 名 と 被験者 10 /学習 8 名)

## 第4章

### 結論

音響シーンの研究の一案として、記述により発生する擦過音である運筆音を用いた文字認識についての検討を行った。この研究により、運筆音という人間にとって雑音とされる音からの情報漏えいの提議などが考えられる。音声認識などは手法が確立されつつあるが、運筆音を対象とした研究は少なく、これによる文字認識はなされていない。本稿では数字0から9という単純な文字を対象として運筆音による文字認識の最適な手法についての検討を行った。

HMM 状態数および特徴量の設定方法について、右利き被験者 10 名の運筆音データを用いた認識実験を行った。HMM 状態数の設定は数字ごとの平均記述時間を求め、これを基に状態数を割り当てた。特徴量の設定に関しては、数種類の特徴量に関する実験を行った。

実験の結果として、数字ごとの平均記述時間より HMM 状態数を割り当てる場合、0.5 s ごとに少なくとも 4 状態が必要であることが確認できた。特徴量に関しては数種類の特徴量を検討した結果、MFCC が最も認識率が高い結果となった。また、運筆音のパワーに対応する 0 次成分や時間変動要素である微分・加速度が運筆音による認識において有効であることも確認した。特に微分・加速度の時間変動要素に関しては曲線を持つ数字の判別精度向上に寄与した。その他に特徴量に用いる次数についての実験を行ったが、数字ごとに認識率の変動が生じるものの、平均認識率には大きな変動は確認できなかった。また、実験を通して学習に用いる運筆音データを増加することで認識率の向上を確認した。

被験者 10 名による認識実験で確認された HMM 状態数および特徴量の最適条件で、収録した全右利き被験者 23 名による認識実験を行った。結果として、本稿の認識条件下における運筆音を用いた手書き数字認識は約 9 割の正答が上限と考えられた。

以上の結果より、運筆音からの手書き数字認識を検討し、約 9 割の正答を得ることが可能であることを確認した。本研究は良好な条件下で行われたものであるが、運筆音による情報漏えいの危険性が示唆されたと考える。

また、本研究の認識では個人差が発生していた。これに関する検討も行い、記述速度がその一因であると考えた。今後の課題としてはこの個人差に対して、記述速度のバリエーションの増加や認識条件の変更等が挙げられる。





## 付録 A

# 本人データのみを使用した認識について

被験者 10 名の運筆音データを用いた認識において、特徴量を Table.3.1、状態数を数字ごとの平均記述時間より 0.5 s で 4 状態とした実験で被験者 H は最も低い認識率を示した。被験者 H の認識で 8 名を学習に用いた際の平均認識率は 74.5 % であった。これは他の被験者と比較して低い数値であり、個人差の発生が考えられた。

また、このように低い認識となった原因の一つとして学習データの使用法があと考えた。被験者 10 名を用いた認識において自己データを認識に用いる際、学習には自己のデータが含まれない。そこで本節では本人の運筆音データのみを使用した場合の認識について検討した。

使用する被験者データはこれまでに個人差が大きいと予測された被験者 H の運筆音データである。被験者 H の運筆音データ、各数字 25 回、合計 250 回を 25 のグループに分割して実験に用いた。1 グループには各数字 1 回、合計 10 回の運筆音データが含まれている。25 グループのうち 24 グループを学習データ、残り 1 グループを認識データとして用いて認識実験を行った。なお、使用した特長量および状態数に関して Table.3.1 および平均記述時間より 0.5 s ごとに 4 状態として行った。

結果として、被験者 H の運筆音データのみを使用した実験の平均認識率は 90.4 % となり、先の結果と比較して約 16 % の向上が確認できた。また、認識結果の詳細について Table.A.1 に示す。Table.A.1 より、数字 0 の認識は低いですが、それ以外の数字では 80 % 以上の正答を得ることができた。また、数字 6 に関しても Fig.3.17 と比較して約 40 % の向上が確認できた。運筆音からの文字認識において、本人の運筆音データ学習の重要性を確認した。

Table.A.1 被験者 H のみを使用した認識の内訳

Target digits	Recognition rate [%]									
	0	1	2	3	4	5	6	7	8	9
0	64	8	0	0	0	0	28	0	0	0
1	0	100	0	0	0	0	0	0	0	0
2	0	0	100	0	0	0	0	0	0	0
3	4	0	0	96	0	0	0	0	0	0
4	0	0	0	0	100	0	0	0	0	0
5	0	0	0	0	0	100	0	0	0	0
6	4	8	0	0	0	0	88	0	0	0
7	0	4	0	0	0	0	0	96	0	0
8	0	0	0	0	0	0	0	0	100	0
9	0	0	0	0	0	0	0	0	0	100

## 参考文献

- [1] Li Zhuang, Feng Zhou and J. D. Tygar “Keyboard Acoustic Emanations Revisited,” TISSEC, volume 13, Issue 1, 2009.
- [2] K. Wuttiwat, 野村 英之, 鎌倉 友男, 上田 浩次, “車のタイヤ音を利用した路面状況の検出,” 信学技報 [応用音響], EA2008-125, 107(470), 59-64, 2008.
- [3] 星野 博之, “ドライバ支援のための走行音による接近車両検知システム,” 日本音響学会誌, 62(3), 265-274, 2006.
- [4] エバンズ 直子, 栗栖 清浩, “車両の自動検出と分類 —都市センシングを目指して—,” 2014 年秋季音響学会, 1539-1540, 2014.
- [5] 岡本 亜紗子, 林田 亘平, 中山 雅人, 西浦 敬信, “多重衝突音のパワー包絡を用いた危険音検出の検討,” 音講論集 (秋), 1551-1552, 2014.
- [6] 前島 隆也, 安川 博, 板井 陽俊, “足音を用いた歩行者数推定に関する一検討,” 信学技報 [スマートインフォメディアシステム], SIS2009-72, 107(470), 129-133, 2010.
- [7] 野村 祐基, 磯村 太郎, 板井 陽俊, 安川 博, “DP マッチングによる歩行足音の個人識別に関する検討,” 信学技報 [信号処理], SIP2007-108, 107(470), 73-77, 2007.
- [8] 庄司 正成, “マイクロホンアレーによる足音位置推定の検討,” 信学技報 [電気音響], EA2009-86, 109(286), 61-66, 2009.
- [9] 井本 桂右, 小野 順貴, 植松 尚, 大室 伸, “イベント遷移を考慮した音響トピックモデルによる欠損を含む観測からの音響シーン推定,” 音講論集 (秋), 1531-1534, 2014.
- [10] 西田 昌史, 北岡 教英, 武田 一哉, “ユビキタスセンシングに基づく日常生活行動データベースの構築,” 音講論集 (秋), 1557-1558, 2014.
- [11] 松山 みのり, 西村 竜一, 河原 英紀, 山田 順之介, 入野 俊夫, “うっかり者を手助けする環境音認識アプリの開発について,” 音講論集 (秋), 1559-1560, 2014.
- [12] 土屋 喬, 小宮山 摂, 武藤 剛, “運筆音を活用した書字訓練装置の開発,” ヒューマンインタフェース学会論文誌, 12(4), 451-457, 2010.
- [13] 金 ジョンヒョン, 橋田 朋子, 大谷 智子, 苗村 健, “筆記音のフィードバックが単純な筆記作業に及ぼす影響の検討,” 日本バーチャルリアリティ学会論文誌, 17(3), 289-292, 2012.
- [14] 牟田 将史, 石川 優, 里井 大輝, 星野 准一, “i 文鎮: 書くことを楽しくする文鎮,” 情処研報 [ヒューマンコンピュータインタラクション], 2014-HCI-157(60), 1-5, 2014.
- [15] Andrew J. Viterbi, “Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm,” IEEE Transactions on Information Theory, 13(2), 260—269, 1967.
- [16] A. P. Dempster, N. M. Laird and D. B. Rubin, “Maximum Likelihood from Incomplete Data

- via the EM Algorithm,” Journal of the Royal Statistical Society, Series B (Methodological), 39 (1), 1—38, 1977.
- [17] L.E.Baum, T.Petrie, G.Soules and N.Weiss, “A Maximization Technique Occurring In The Statistical Analysis of Probablistic Functions of Markov Chains,” The Annals of Mathematical Statistics, 41(1), pp.164-171, 1970.
- [18] “数字の練習プリント,” <http://happyililac.net/suuzi.html>, (2015/1/28 アクセス)
- [19] “HTK Speech Recognition Toolkit,” <http://htk.eng.cam.ac.uk/>, (2015/1/28 アクセス)
- [20] “WavePad 音声編集ソフト,” <http://www.nch.com.au/wavepad/jp/>, (2015/1/28 アクセス)

## 謝辞

本研究を進めるにあたり，終始適切なご指導とご助言をいただきました，本学工学部准教授 西野 隆典 先生に深く感謝します．本研究に対し，貴重な助言を頂きました，本学工学部 教授 成瀬 央 先生に厚くお礼申しあげます．また，学生生活を共にした，パターン情報処理研究室の皆様感謝します．





## 本論文に関する著者の発表文献

1. 北野 真示, 西野 隆典, 成瀬 央, “運筆音による手書き数字認識についての検討,” 平成 25 年度電気関係学会東海支部連合大会講演論文集, F2-5, 2013.
2. 北野 真示, 西野 隆典, 成瀬 央, “HMM を用いた運筆音からの手書き数字認識,” PRMU 2013-88, 121-125, 2013.
3. 北野 真示, 西野 隆典, 成瀬 央, “運筆音による手書き数字認識における学習データ数の調査,” 平成 26 年度電気・電子・情報関係学会東海支部連合大会講演論文集, K5-3, 2014.
4. 北野 真示, 西野 隆典, 成瀬 央, “運筆音を利用した手書き数字認識,” 2014 年秋季音響学会, 3-8-9, 1545-1546, 2014.
5. Shinji Kitano, Takanori Nishino, and Hiroshi Naruse, “Handwriting Numeric Character Recognition Using Writing Sound,” Proc.IS2EMU2014, B-16, 2014.
6. 北野 真示, 西野 隆典, 成瀬 央, “運筆音による手書き数字認識の特徴量の調査,” 2015 年春季音響学会, 1-Q-23, (2015.3 発表予定)