



Title	顔の情報処理機構 : 視覚的イメージの形成過程
Author(s)	乾, 敏郎; 宮本, 健作
Citation	大阪大学人間科学部紀要. 1979, 5, p. 191-221
Version Type	VoR
URL	<a href="https://doi.org/10.18910/12045">https://doi.org/10.18910/12045</a>
rights	
Note	

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

# 顔の情報処理機構

—視覚的イメージの形成過程—

行動工学 乾 敏 郎  
宮 本 健 作

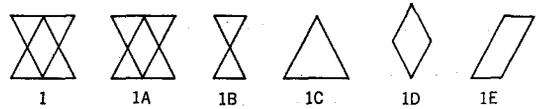
## 顔の情報処理機構

### —視覚的イメージの形成過程—

#### 1. 緒言

外界を注視したとき、網膜に投影された像は、並列にある程度の前処理が施された後 (Von Wright, 1970), ICON (もしくは VIS; Visual Information Storage) と呼ばれるバッファに貯えられる (Sperling, 1960)。

ICON の内容を基礎にして、視覚的に統合された内部表象を視覚的イメージ (visual image) と呼ばれている。<sup>1)</sup> 視覚的イメージは決して写真のようなものではなく、知覚体制化の法則に従って分節が生じている。また、ICON の内容は注視点が変わるごとに書きかえられるが、視覚的イメージはさらに長い時間持続し、 $\frac{1}{4}$  sec で制限される事はない。<sup>2)</sup> それゆえ、視覚的イメージは各注視点で



の部分情報が統合されたものであり、この意味で視覚的世界全体の表現 (Haber と Hershenson, 1973) である。このような視覚的イメージを A. M. Farley (197

図1 パターンを写真のように記憶しているのではなく、部分間の関係によって構造を表現している。いわゆる埋没図形に関する実験は、この事実を端的に示す。1に示す図を被験者に提示した後、1A~1Eの図形を見せ、それが1の部分であったかどうかを答えさせる。1Eを正答した人は、ごく少数であった。(Reed, 1973による。)

6) は、次のように定義している。「視覚的イメージとは、知覚時に視覚的環境に接近する内的過程と同型な過程によって直接再生可能なように構成されたものである。その視覚的イメージは、視覚的特徴および空間的關係の内的意味論的表現であるが、十分に外的事象に対応しているとき、それは知覚に分類される。」(Farley, p479) 視覚的イメージが貯えられている状態の記憶を Visual STM (Short-term Memory) と呼んでいる。従来、記憶に関する研究は一般に言語的なもの (phonemes, syllables, words, semantic primitives) が多く、最近ようやく視覚的イメージの構成単位やイメージ形成のアルゴリズムの研究が行なわれるようになった。Haber と Hershenson (1973) は、その著書「視知覚の心理学」でこのような事情を次のように述べている。

「不幸にも研究室で行なわれる多くの知覚実験では、知覚者はただ一回の注視が可能な時間しか与えられていない。それゆえ、ICON の特性と視覚的イメージの特性とを区別することが困難である。…… (中略) ……いかにしてこれらの統合と構成がなされるのかという

点について、私の考えを導いてくれる証拠はほとんどない。事実、Gibson や Hochberg などごく少数の研究者を除いては、ほとんどこの過程がどのようにして起こるのかということに注意を払っていない。」(Haber と Hershenson, p174)

視覚的イメージの形成に関する研究を困難にしているのは、その過程がきわめて**能動的**であり、かつ**継時的**であるということである。ここで、視覚的イメージの形成過程について概略をまとめてみよう。私たちは、与えられた画像を一様に走査 (scan) するのではなく、またランダムに走査するでもない。私たちは、適当なアルゴリズムにしたがって、画像の部分を「選択、して走査している。この**選択性 (もしくはサンプリング)**が、この過程の第一に重要な要因である。これは、ある領域に対して加重操作を施すことに他ならない。これはまた、注視領域以外の情報が上位中枢に伝達されるのを抑制することに対応する (Inui, et al, 1978)。第2に重要な点は、これらの選択が**継時的に**起って、**全体像が形成される**ということである。第3に重要な点は、**ごく短時間で重要と思われない領域の却下**が起こるということである。Mackworth と Morandi (1967) は、一枚の絵を  $8 \times 8$  のマトリクス状に分割し、第1の被験者群には、それぞれの画素に対する情報の度合いを10段階に評定させた。また第2の被験者群には、絵全体を視角にして  $16^\circ \times 16^\circ$  の大きさで10秒間提示し、このときの眼球運動を測定した。その結果、第1群の評定から Mackworth が、情報集中領域 (informative area) と考えた領域に注視点が1回目か2回目の注視で移動すること、情報量に関する評定と注視回数との間にきわめて高い相関をもつことを見出している。この機能は、明らかに周辺視によるものであり、周辺視は、規則性を乱すもの、ならびに高次中枢で立てられた各時点における仮説に符号しない領域を検出する能力があると考えられる。私たちは、各部分を注視するとき、それまでの注視点から得られた情報と、今注視している部分からの情報をもとにして、画像全体の**仮説を強化する**。そして、さらにその**仮説を確認**するために、その時点における新たな「情報集中領域」を注視しようとする。このように、私たちの視覚系を含むより複雑な認知システムは、たえず「**仮説の生起と検証**」 (hypothesis generation & verification) をくりかえしながら、画像全体を把握しようとする。この点について別の見方をすれば、**bottom-up** (部分から全体へ) ならびに **top-down** (全体から部分へ) という情報処理様式をくりかえしながら、全体を把握しようとしているとも考えられる。

以上を要約すると視覚的イメージの形式過程においては、画像の各部分が継時的に選択され、かつ統合されていく。さらに各部分の選択は、仮説の生成と検証という、言わば下から上への情報の流れと、上から下への情報の流れのサイクルの中で行なわれる。この事実を最も簡単に示してくれるのは、図2に示す不可能図形 (impossible figure) である。ある部分を注視し、全体に対する仮説をもうけてその仮説にしたがって、他の部分の処理をしようとする。これを検証するために、他の部分に注視点を移動させる。このような上から下への処

理のために一瞬“意味のある物、として  
認知されるが、つぎの瞬間には、その仮  
説はくずれてしまい新しい仮説が生起す  
るのである。

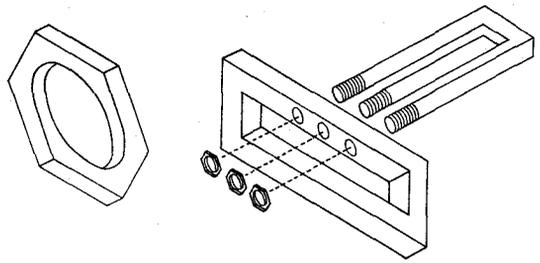


図2 不可能図形 (Raphael, 1976より)

さて、認知心理学のテーマは、網膜の  
種々の部分に投影された情報を、いかに

して統合し、一つの応答を示すのかとい  
う問題を明らかにすることである。したがって、上述の視覚的イメージの形成過程のアルゴ  
リズムを考えることは、きわめて重要であると思われる。しかし、現実の日常生活におい  
て、私たちが見るのは意味あるパターンであり、無意味な図形や文字列ではない。そこで、  
刺激パターンとして何を採用するかが、問題となる。対象の選択のしかたいかんでは、異な  
る結果が生じる可能性がある。たとえば精神医  
学の分野で、失認症に関する研究があるが、表  
Iに示すごとく、物体失認、相貌失認、視覚失  
認性失読、同時失認など対象に限定された失認  
が生じる。それゆえ、私たちは、物体、文字、  
文章、風景、顔等、異なる対象に関連したパタ  
ーン認識のメカニズムについて、種々の異なる  
面から着実にデータを集積することによって、  
はじめてヒトのパターン認識過程が明らかにされてくると考えられる。

表 I

認 識	視 覚 失 認 症
物 体 認 識	物 体 失 認
相 貌 認 識	相 貌 失 認
文 字、単語の認識	視 覚 失 認 性 失 読
絵、風景の認識	同 時 失 認

本論文では、顔の認識という課題を通じて視覚的イメージの形式過程について考察する。  
このような課題を選んだのは、主として次の3つの理由による。第1に、人の顔を見るとき、  
全体的印象が先行し、顔の部分は知覚されないことがあると、古くから言われていて、  
画像の中でも特殊なものであること。<sup>3)</sup>

第2に、部分的な構成要素それ自体で一つの意味のある画像であるが、全体として一つの  
概念的にまとまった画像であること（言いかえると、一つの知覚ゲシュタルトを構成する。  
Miller (1956) の言葉を使うならば、単一のチャンクが存在するという事になる。）。第3  
は、仮に知っている人の顔であれば、その人の名前を言わせることで、被験者が全体を把握  
できたかどうかを知ることができるということである。顔の一部の特徴だけを提示した場合  
には、それが誰の顔であるのか言いあてることは困難である (Appendix 参照)。

これらの理由によって、本論文では、人の顔の認識という課題を通して視覚的イメージの  
形成過程を考察する。

まず、次章では、どの程度の視野の広さからどのような情報を取り入れているのかという

問題について述べ、次いで第3章では主として、情報処理の時間経過を論じることにした。

## 2. 顔の認識における視野の役割

眼球運動測定を通して、パターン認識の研究が行なわれるようになって以来、すでに数十年を経過しようとしている。初期の研究として、Sisson (1935) による反転図形と眼球運動の関係を調べたもの、Brandt (1940 a, b) による注視点の視野左上部への偏りを報告したものの、また比較的最近のものとして、鳥居ら (1963) による錯視と眼球運動を調べたものなどがあり、それぞれ一応の成果をみることができる。しかし、すべての注視点で同じ情報処理が施されているのかどうかという問題については、十分には明らかにされていない。また、注視点の周囲のどの範囲の処理がなされているのか全くわかっていないことが、こうした研究のもつ残された課題であった。近年、一部の工学者によって、パターン認識における視野の広さの効果が、検討された。たとえば、Watanabe と Yoshida (1973) は、大きさが  $15^{\circ}$  -  $17^{\circ}$  のパターンを中心視 ( $3^{\circ}$ ) のみで見た場合、その形は把握されても、パターンのもつ全体の意味は了解されないことを報告している。全体のパターンと視野の広さの割合が、パターン認識にとって重要であることは、斎田らの報告 (1975, 1976) にも示されており、視野の広さが対象全体の半分以下になると、認識能力が視野の広さに応じて低下すると述べている。ここでいう視野とは、視野計で測定されたものではなく、パターン識別が可能な範囲のことで、いわゆる有効視野 (池田, 1975) と呼ばれるものを意味する。本章では、中心視で扱えられる視角  $5^{\circ}$  程度の大きさの顔の認識における視野の役割をいくつかの条件下で述べた実験成績について述べる。

### 実験の遂行に関する留意点

実験の基本的な課題は、誰の顔かを判断することであるから、実験には、きわめて知名度の高い有名人の顔 (public face) 写真が用いられた。実験に際して、以下の点に留意した。

#### i) 個人差

上述のごとく、視覚刺激として有名人の顔写真を用いたが、各被験者によって顔写真に対する知名度が異なることが考えられる。そこで、実験終了後、以下に述べる知名度テストを行ない、その結果、認識可能であった顔写真に対する実験結果のみをデータとして採用した。

知名度テスト ; 実験に使用した顔写真全体を TV ディスプレーに 80 msec 間提示し、誰の顔かを判断させた。その提示時間の前後は、常にノイズマスクが提示された。このテストによって、個々の被験者についての写真に対する知名度がわかる。提示時間は、予備実験の

結果から決定した。

### ii) 顔写真の差異

たとえよく知られた顔写真であっても、見やすさの点で差が生ずる。そこで、本実験で使用する顔写真120枚を、知名度が偏らないよう配慮しながら、20枚ずつ6グループに分けた。基本的には各実験条件にこのうちの1グループを用いた。各条件に用いられる写真グループは、被験者ごとに変えることによって、顔写真の見やすさや、知名度の差からくる影響を極力なくすように注意した。一例として、10名の被験者による知名度テストの結果は、表Ⅱに示すごとくである。グループ間での正答率の差は小さいことから、顔写真が平均して分類されていることを示している。

### iii) 被験者への教示

実験の開始に際して、被験者に与える教示は、最小限にとどめた。「これから、目の前にあるテレビのディスプレイに、有名人の顔が出ます。それらは、歌手、俳優、落語家、政治家、野球選手です。」

### iV) グループ間の差

一度、実験に参加した被験者は、どんな顔写真が出るのかわ知っているため、再び実験に参加することはできないので、各実験ごとに被験者グループを変えた。グループ間の差がないことを確かめるため、各実験ごとに共通のパラメータを設定して、その値をもとにグループ間の差の検定を行なった。

### V) 実験者効果

被験者と同室して、記録をとる実験助手（アルバイトの女子学生）は、実験内容および目的を知らされていなかった。実験者らは実験室の外にいて、実験の遂行と装置の調整にあたった。また、被験者には、正誤の結果は知らせなかった。

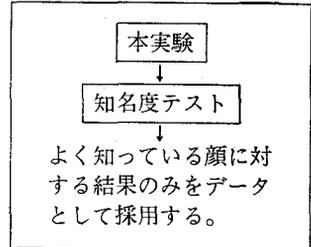


表 Ⅱ

グループ	正答率
1	74.5%
2	70.5
3	76.0
4	70.0
5	74.0
6	73.0

## 2.1 実験Ⅰ：視野制限と認識能力との関係

被験者が同時に見える視野の広さと、認識能力との関係をしらべる。

### 2.1.1 方法

装置…図3に示すようなテレビジョンミキシングシステムを使用した。被験者は手もとのレバーを上下左右に動かすことにより、テレビ上で視野を自由に移動させることが可能である。

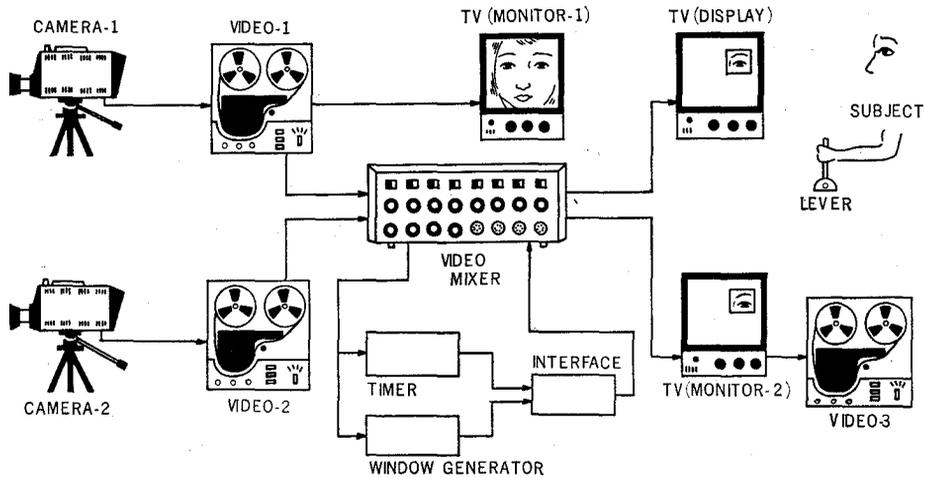


図3 実験システムの概略図。カメラ（もしくはビデオ）1で撮った顔写真全体の映像と、カメラ（もしくはビデオ）2で撮った背景の映像と、視野の位置と大きさを決めるデジタル回路の出力をテレビミキサーで合成する。背景の映像は、実験4の時だけ使用した。被験者は、手元のレバーを動かして、視野の位置を自由に動かすことができる。実験者は、別室でモニターテレビによって、被験者の走査過程を記録、観察した。

刺激…対象として、有名人の顔写真120枚を20枚ごとの6グループに分けて用いた。各グループごとに、視野の広さを変化させた。視野の広さは、視角にして $0.9^{\circ} \times 0.9^{\circ}$ 、 $1.3^{\circ} \times 1.3^{\circ}$ 、 $1.7^{\circ} \times 1.7^{\circ}$ 、 $2.3^{\circ} \times 2.3^{\circ}$ 、 $3^{\circ} \times 3^{\circ}$ 、 $3.8^{\circ} \times 3.8^{\circ}$ の6種類とした。被験者とディスプレイの距離は、155cm。テレビは $5.3^{\circ} \times 5.3^{\circ}$ の正方形の穴が空いている板で覆われており、顔写真は、その正方形一ばいに提示されるようにカメラで調整されているが、被験者からは窓（図3に示されたTVディスプレイの左目のみが見えている四角）の部分しか見えない。

手続…スタート合図とともに、被験者は手元のレバーによって、テレビディスプレイ上の視野の位置を動かして、顔写真を見る。このとき、被験者がその顔写真が誰であるかがわかるまでに要する時間を測定した。ただし各条件によって、あらかじめ制限時間をもうけ、その時間内での認否をしらべ、かつ反応時間を測定した。その制限時間は、視角が $0.9^{\circ} \times 0.9^{\circ}$ のときは90sec、 $1.3^{\circ} \times 1.3^{\circ}$ 及び $1.7^{\circ} \times 1.7^{\circ}$ の場合は60sec、 $2.3^{\circ} \times 2.3^{\circ}$ では20sec、 $3^{\circ} \times 3^{\circ}$ および $3.8^{\circ} \times 3.8^{\circ}$ は15secとした。被験者の課題は、提示された顔が誰であるかを判定することである。しかし、たとえ名前が思い出せない場合でも、その人が何をやる人か、あるいはどんな事件に関係がある人かを答えることによって、実験者に被験者の解答が誰を意味しているかがわかればよい。（名前以外の表現を認める。）実験終了後、知名度テストを行なった。

被験者…男子3名ならびに女子7名（平均年齢22才）で、全員視力は正常であった。

## 2.2.1 結果

知名度テストの結果, 判定可能でかつ, 正解の判定所要時間から, 平均反応時間を求め, 視野面積との関係を調べた。被験者10名の平均値を図4に示した。図中の▲印は, 顔写真全体が提示されてから反応するまでの平均反応時間で, これは別の実験で得られたデータである。図4に示されるごとく, 視野を次第に制限して, 全体の18%まで縮小されると, 正答率(黒四角: ■印)は, 有意に低下し( $\chi^2=11.9$ ,  $p<0.01$ ), 誤答率が増加した。また, 視野が全体の5%になると, 正答率は50%に低下し, さらに3%以下になるときは, ほとんど認識不可能であった。この点については, 予備実験において, 視野の広さが $0.75^\circ \times 0.75^\circ$ などの場合について調べた。図4の黒丸(●印)は, 正答時の平均反応時間を示している。視野の広さのもつ有意効果が検出された( $F(5,592)=51.2$ ;  $p<0.01$ )。

図5は, 正答率で重みづけられた平均反応時間と視野の広さとの関係を示している。視野が全体の18%以下になると, 反応時間の急激な増加が見られた。

次に, 図6は誤判断の割合(誤判断の回数を, 判定不能の場合も含めた全誤答数で割ったもの)とその平均反応時間を示している。視野が全体の18%より狭くなると有意に誤答(●印)が増加し, さらに10%より狭くなると, 誤答率が低下した。これは10%以下の場合には, 認識不可能な回数が急激に増加する事を意味する。一方, 反応時間(▲印)は視野の広さの単調関数になっている。

顔の認識は, 局所的な手がかりを総合して認識しているように思われる。しかし, 認識するためには, 視野の広さがある一定の大きさより広くなければならないことが明らかにされ

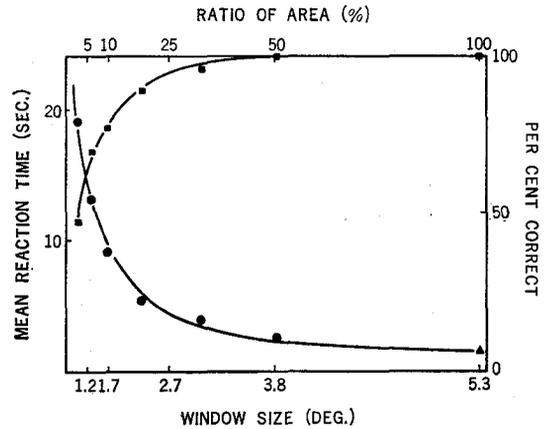


図4 平均の正答率と視野の広さとの関係(●)および, 正答時の平均反応時間と視野の広さとの関係(●)。視野が全体の18%にまで狭くなると, 有意に正答率は低下した。又, 正答時の平均反応時間も, 視野が18%より狭くなると急激な増加を示した。

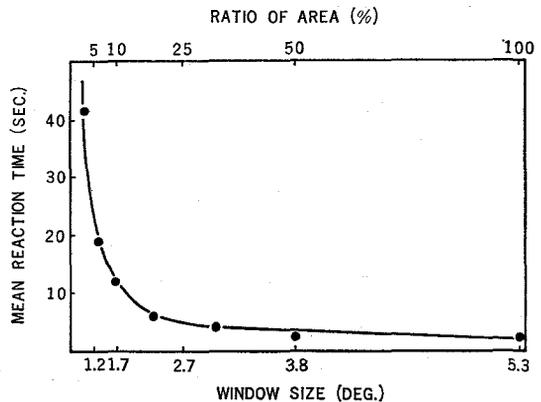


図5 正答率で重みづけられた平均反応時間と, 視野の広さの関係。縦軸は, 図4の平均反応時間(図4の丸印)と正答率(図4の四角印)をかけた値である。正答率で重みづけられた平均反応時間は, 視野の広さが18%以下で急激な増加が見られた。

た。

### 2.2 実験Ⅱ：ボケの度合と認識能力

実験Ⅲを行なうための予備データを得る目的で施行した。2次元 low-pass-filter の遮断周波数と認識能力との関係をしらべた。

#### 2.2.1 方法

装置…実験Ⅰと同じテレビミキシングシステム(図3)を用いた。

刺激…実験Ⅰで用いた6グループの顔写真のうち、5グループ100枚を用いた(提示されたときの顔の大きさは、ほぼ一定)。5.3°×5.3°のディスプレイに矩形波格

子(grating)を提示しこれが完全に見えなくなるまで、カメラのピントをボカす。この状態で顔写真を提示する。矩形波の空間周波数は、20c/display, 14c/display, 10c/display, 8c/display, 6c/display の5種で、これらは3.8cpd(cycle per degree), 2.7cpd, 1.9cpd, 1.5cpd, 1.2cpd に対応する。この空間周波数の値をここで、遮断周波数(cut-off frequency)と呼ぶことにする。

手続…スタートの合図とともに、顔写真がディスプレイ全体に5sec間提示される。被験者にその顔が誰であるかを答えさせた。提示開始後5secを過ぎてから答えてもよい。この場合も実験Ⅰと同様、名前以外の表現も許した。

被験者…男子3名、女子3名(平均年齢20才)で、全員視力は正常であった。

#### 2.2.2 結果

まず、正答率と遮断周波数との関係を図7に示した。視力1.0の者は、高いコントラストをもつ150c/displayの矩形波を識別することが可能である。図7にみられるごとく、高周波成分を削ってゆき、遮断周波数が20c/displayになってもまだ100%認識可能であった遮断周波数が9c/displayのときには、50%の正答率が得られた。遮断周波数が14c/display以下の顔写真では3cpd以下の低周波数成分のみで認識していることになる。この事実は、視覚系の空間周波数チャンネルの理論(この点については、先の論文、乾(1978)を参照された

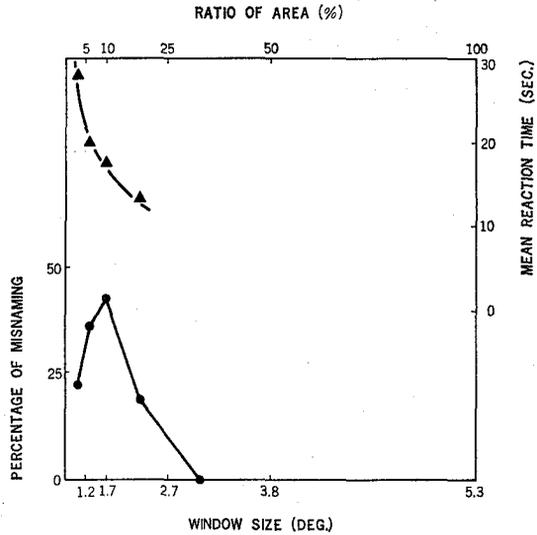


図6 誤判断の割合(誤判断の回数を、判定不能時も含めた全誤答数で割った値)と視野の広さの関係(●), および誤判断時の平均反応時間と視野の広さの関係(▲)を示す。

い)との対応を考えることは興味ぶかい。

Harmon (1973)も述べているように、予想外に人間の認識能力が高かった。局所的な手がかりは、遮断周波数が12~13 c/display の付近で全くなくなる。この場合の局所の手がかりとは、目や鼻の形などの部分情報を意味する。

### 2.3 実験Ⅲ：局所情報の欠如時の視野の影響

実験Ⅱの成績で明らかのごとく、私たちはたとえ目や口の形といった局所的な情報がない状態においても、人の顔を認識することができる。このことは、全体的なイメージとも言うべきものを手がかり

りとして、顔を認識していることを示唆している。そこで、本実験においては、全体的な顔の情報処理機構の手がかりを得るために、実験Ⅰと同様な方法によって、視野の広さと認識能力との関係をしらべた。

#### 2.3.1 方法

装置…実験Ⅰと同じテレビミキシングシステム(図3)を用いた。

刺激…実験Ⅱで使用した刺激のうち、10c/displayの遮断周波数のものを使用した。その理由は、実験Ⅱの条件のうち、局所的情報が全くなく、かつ比較的正答率が高かった(70%)ことによる。この条件で提示した写真を、実験Ⅰと同様、視野を制限して認識させた。この場合、視野の広さは $2.3^\circ \times 2.3^\circ$ 、 $3^\circ \times 3^\circ$ 、 $3.8^\circ \times 3.8^\circ$ 、 $5.3^\circ \times 5.3^\circ$ の4種とした。

手続…スタートの合図と共に、前述のごとく、被験者は窓の位置をレバーによって動かしながら、顔写真を見た。このとき、被験者が誰の顔写真であるかわかるまでの時間を測定した。ただし、各条件によってあらかじめ制限時間をもうけ、その範囲で認否をしらべ、かつ認知時間を測定した。制限時間は、 $2.3^\circ \times 2.3^\circ$ と $3^\circ \times 3^\circ$ が30 sec、 $3.8^\circ \times 3.8^\circ$ が20sec、 $5.3^\circ \times 5.3^\circ$ が5secとした被験者の課題は、提示された顔が誰であるかを判定することであり、実験Ⅰ、Ⅱと同様、名前以外の表現を認め

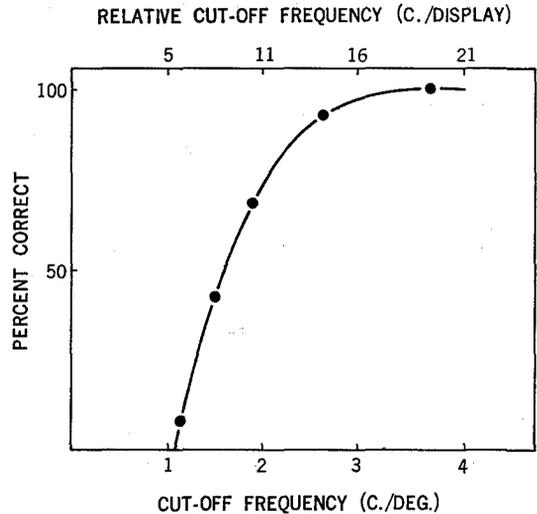


図7 低域遮断周波数と正答率。遮断周波数が12~13c/displayより低くなると、局所の手がかりは全くなくなる。しかし、この状態においても、かなり高い正答率が得られた。

被験者…男子3名および女子3名（平均年齢21才）で、全員視力は正常であった。

2.3.3 結果

図8は、正答率（■印）と正答時の平均反応時間（●印）を示している。全体を同時に見せた場合の平均正答率は75%であり、これは実験Ⅱの70%とほぼ一致していた。したがって、実験Ⅱの実験者群と本実験者群との間には、大きな差はないと考えられる。視野の広さを制限してゆき、その広さが全体の33%になると、正答率（■印）は、有意に低下した（ $\chi^2=16.02$ ;  $p<0.01$ ）。また、平均反応時間には、視野の広さの有意効果が検出された（ $F(3,123)=29.8$ ;  $p<0.01$ ）。さらに、全体を同時に見せた場合の正答率を、100%に正規化して50%閾をしらべると、それは視野の広さが全体の25%であることがわかった。

図9に、正答率で重みづけられた平均反応時間と視野の広さの関係を示す。この場合も、視野が33%と50%の間から、急激に反応時間の増加が見られた。

2.4 実験Ⅳ：背景情報の効果

私たちが外界にある物を見るとき、注視している部分の周辺には、全く情報がないわけではない。周辺からは、かなりボケた情報をとり込んでいるように思われる。本実験では、ほとんどそれだけではパターン認識ができない情報でも、それが背景にあった方が認識しやすくなるのかどうかをしらべる。

2.4.1 方法

装置…実験Ⅰに述べたものと同じ装置を用いた。

刺激…実験Ⅰで用いた写真のうち、2グループ分、計40枚の写真を使用した。これらの写

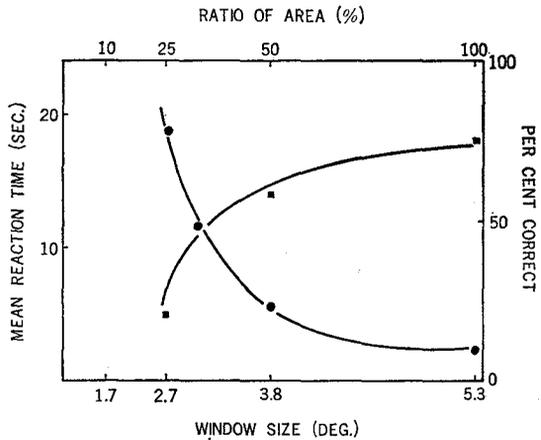


図8 平均の正答率と視野の広さの関係（■），および正答時の平均反応時間と視野の広さの関係（●）。

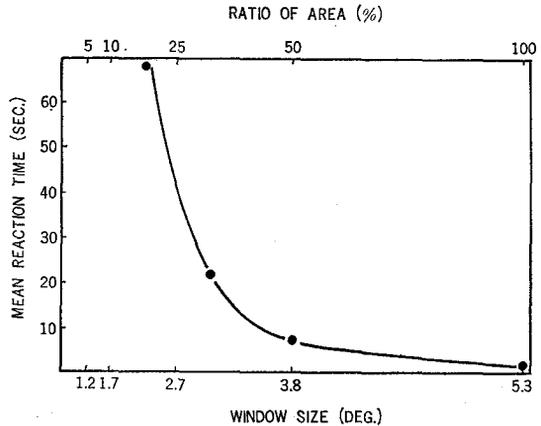


図9 正答率で重みづけられた平均反応時間と、視野の広さの関係。縦軸は、図8の平均反応時間（●）と正答率（■）を掛け合わせた値である。正答率で重みづけられた平均反応時間は、視野の広さが、50~33%の間から急な増加が見られた。

真を、実験Ⅰの場合と同様、視野を制限して認識させた。視野の広さは、 $1.3^{\circ} \times 1.3^{\circ}$ とした。2グループのうち1グループは、実験Ⅰと同様に $1.3^{\circ} \times 1.3^{\circ}$ の視野以外の部分は全く情報がない条件で走査させた。他の1グループは、制限視野以外の部分にいわゆる背景情報としてボケた画像を入れた。背景情報は、実験Ⅱで用いた顔写真のうち、遮断周波数が6c/displayのものを用いた。この場合、図3に示したごとく、1台のカメラにピントを正常に合わせた状態で、他のカメラは、実験Ⅱで述べた方法によってピントをぼかして顔写真を撮った。

手続…スタートの合図とともに、前述の実験と同様に、被験者はレバーを操作することによって、窓の位置を随意に動かしながら顔写真を見た。このとき、被験者がその顔写真が誰であるかわかるまでの時間を上述の2条件について測定した。

被験者…男子2名、女子6名（平均年齢22才）で、視力は全員正常であった。

#### 2.4.5 結果

条件別に平均反応時間を表Ⅲに示した。実験Ⅱの結果から、もし背景に用いた顔写真だけなら、わずか8%という正答率しか得られない。このような情報であっても、背景にあった方がない場合に比べて、はるかに効果があった( $p < 0.01$ )。なお、本実験の背景情報がない場合は、実験Ⅰと同じ条件である。両者の結果は、ほぼ一致しており、被験者グループによる差はない事がわかった。

表Ⅲ  
WITH BACKGROUND      WITHOUT BACKGROUND

	WITH BACKGROUND	WITHOUT BACKGROUND
M.R.T. (SEC.)	8.4	12.8

#### 2.5 考察（実験Ⅰ～Ⅳ）

実験Ⅰ～Ⅳまでの結果を要約すると次の通りである。

- (1) 視野が全体の5%以下に縮小されると、顔の認識はきわめて困難であった（正答率50%以下）。
- (2) 視野が全体の5～20%になると、全体のイメージをつかむことはできないが、局所情報の累積によって、やや不完全ながら誰の顔かを認めることができた。
- (3) 視野が全体の33～50%の広さで、ようやく全体のイメージがつかめるようになった。
- (4) ボケた情報（それだけではほとんど認識できない）を背景に入れた場合、認識速度は背景がないときに比べて、有意に速くなった。

これらを総合すると、視覚的イメージと記憶との照合には、同時に見なければならぬ最小面積が必要であり、それより視野が小さい場合には、いくら継時的に走査しても記憶との照合ができない。このような事実は、記憶との照合過程の背景には最小単位が、不可欠なもの

として存在することを想像させる。しかしながら、逆に視野がある大きさを越えると、もはや視野の大きさは認識速度に対して有効にきかなくなる。この事実を考慮すると、有効視野には上限が存在することが考えられる。<sup>4)</sup> ほぼ正確に顔を認識するためには、局所情報がある場合は18%以上の視野が必要であり、全体的なイメージによる場合には33%より大きな視野が必要であった。これに関連して、Raynerらは、読書時の視野の役割についてしらべた一連の実験から、次のような結論を得ている。(i) 注視点より右12~15文字の範囲からは、単語の長さの情報を処理し、注視点より右10文字までの範囲からは、単語の全体的な形の情報を処理している (McConkie と Rayner, 1975)。(ii) 注視点より右1~6文字の範囲内で、単語の意味内容の情報処理をしている (Rayner, 1975)。Raynerらによる文章の一次元走査に関する研究による結論と、二次元走査についての本実験から得られた結論との比較検討を試みることは興味深い。

### 3. 顔の認識における情報処理の時間的要因

Neisser (1967) は、"認知心理学、の中で、人間の情報処理が、大域的な並列処理である前注意過程 (preattentive process) と、直列処理である集中的注意の過程 (focal attentional process) の2つに分けられると述べた。前者においては、全体的なあらい処理がなされ、後者においては、前者の出力を詳細に分析する。Hoffman (1975) や Inui ら (1978) は、これらの2つの過程が継時的に働くことを検証した。さらに、Inui ら (1978) は、集中的注意の過程で生ずる内的走査を *mental scanning* と呼び、その神経生理学的モデルを提案した。本章では、顔の認識における情報処理の時間経過を吟味するとともに、局所的な情報と全体的な情報とが、完全なイメージを形成するための条件について検討する。

#### 3.1 実験V：局所的情報処理と全体的情報処理の時間特性

中心視だけで把えられる程度の大きさで、顔写真を瞬間提示し、顔の種々の特徴の情報処理の手順をしらべる。

##### 3.1.1 方法

装置…実験Iに示したテレビミキシングシステムを用いた。一方のカメラで顔写真をうつし、他方のカメラでノイズマスクをうつした。テレビの1フレームが16msecであるため、これらの2台のカメラは、16msecの整数倍の時間で切りかえられた。

刺激…実験Iで用いた知名度の高い顔写真50枚を使った。顔の大きさは約5°で、これは中心視で把えられる大きさである。一方、ノイズマスクはLETRASET No.416 Screen-Toneを使用した。

手続…被験者に提示された顔の特徴（表IVに示した項目について）を完全に憶えさせた後、各項目についての特徴をいわせた。すべての項目について、表IVに示された表現のみを許した。たとえば、目、については「目が大きい、等である。スタートの合図で、顔

表 IV

		1	2	3	4
写真が提示さ	FACE	ROUND	MEDIUM	NARROW	UNKNOWN
れた。提示時	HAIR	LONG	AVERAGE	SHORT	UNKNOWN
間は、16, 32,	SEX	MALE	-----	FEMALE	UNKNOWN
48, 64, 80m	AGE	TEENS	TWENTIES	THIRTIES	UNKNOWN
sec の 5 種	EYE	WIDE	MEDIUM	NARROW	UNKNOWN
で、顔写真提	MOUTH	LARGE	MEDIUM	SMALL	UNKNOWN
示前後は常に	NOSE	LONG (LARGE)	MEDIUM	SHORT (SMALL)	UNKNOWN

ノイズマスクが示された。顔写真提示後ただちに被験者の報告を求め、15sec 以内には全報告を聞き終った。

被験者…男子 3 名, 女子 3 名 (平均年齢 23 才) で、視力は全員正常であった。

### 3.1.2 結果

まず、図10は見えなかったと答えた割合と、提示時間との関係を各項目別にプロットした結果を示している。この図から、50% 閾 (すなわち 50% 見えなかったと答えた提示時間) を求めた結果を表 V に示す。性別や髪型や顔の形といった全体情報は、およそ 30msec 以前に 50% 閾に達する。一方、目や口や鼻の形といった局所情報は、全体情報よりずっとおくれて 50% 閾に達することがわかる。さらに、目は見えなが

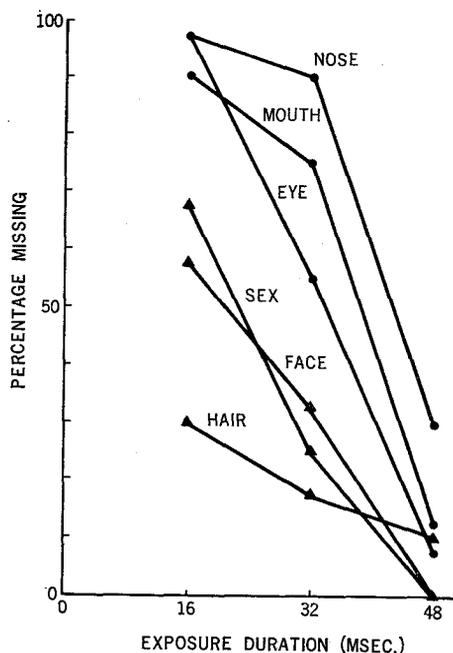


図10 顔写真の提示時間と、評定不能率を各項目別にプロットした。

50% THRESHOLD

FACE	0 < t < 16
HAIR	21
SEX	23
EYE	34
MOUTH	39
NOSE	43

(MSEC.)

表 V

口、鼻は見えなかったとか、目と口は見えたと鼻は見えなかったというように、目→口→鼻という順序で見ている場合が、比較的高い頻度で示された事は注目される。このように、

- (1) 全体情報は局所情報よりも早く処理される。
- (2) mental scanning には、ある程度の順序が決まっているように思われる。(目→口→鼻という順序の傾向が高いと想像される。)

3.1.3 検討

(a) 見えたとか、あるいは見えなかったかという判断基準をより明確にさせるため、各項目別に主観評価を別のグループにやらせた。本実験とは別の3名に、十分長い時間それぞれの顔写真を見せた後、本実験と同じ評価を行なわせた。それを基準にして、本実験の被験者による評価のチェックを行なった。(たとえば3人の評価で、目が普通または大きいとなっているときに、本実験の被験者がそれを小さい

と答えたときは、誤りとした。)このようにして、見えなかった場合も含めて、全体を100%とし正解の割合を表VIに示した。この表に明らかなごとく、“見え

表 VI

EXPOSURE TIME	SEX	AGE	FACE	HAIR	EYE	MOUTH	NOSE
32	47.5 (25)	47.5 (42.5)	65 (17.5)	52.5 (32.5)	30 (55)	17.5 (75)	10 (90)
16	25 (67.5)	12.5 (80)	45 (30)	40 (57.5)	5 (97.5)	2.5 (90)	0 (97.5)

単位は%, ( )内は unknown と答えた割合である。

表 VII

16 msec.							32 msec.						
	SEX	FACE	HAIR	EYE	NOSE	MOUTH		SEX	FACE	HAIR	EYE	NOSE	MOUTH
SEX	—	*		*	*	*	SEX	—			*	*	*
FACE		—		*	*	*	FACE		—		*	*	*
HAIR			—	*	*	*	HAIR			—	*	*	*
EYE				—			EYE				—	*	*
NOSE					—		NOSE					—	*
MOUTH						—	MOUTH						—

\*P<0.01

た、場合は、かなり確かな判断を下していることがわかる。

(b) 提示時間が 16msec, 32msec の場合について、各項目の見えなかった割合の差を検定した(表VII)。16msec では、局所情報と全体情報との間に大きな差が見られた (p<0.01)。32msec になると、局所情報相互間でも、有意な差が見られるようになった。

3.2 実験VI：視野制限の条件下における注視頻度と注視順序—その1—

視野を制限した条件下で、顔を認識する場合の注視部分の頻度および注視順序をしらべた。

## 3.2.1 方法

装置，刺激および手続…実験Ⅰと同様であった。

視野が狭い場合(0.9°×0.9°)について，被験者が走査している様子をビデオテープに記録し，後に再生し解析した。

## 3.2.2 結果

任意に8試行を選び，走査の軌跡を分析した。走査の状態は，2つの成分すなわち，速い部分と遅い部分に分けられるが，特に遅い部分は，眼球運動の注視点に対応していると考えられる。図11は，注視頻度を項目別に示したもので，右目と左目を合計すると，46.7%で最も高く，次いで，口の14%，鼻の9.6%の順に低くなっていく。(まゆ毛の注視頻度が高いのは，視野が狭いため目とまゆ毛が同時に見られないことによる。) つぎに，各顔写真について走査した回数を図12に示した。この場合の走査順序は，左目→右目→口→鼻の順が最も多く

見られた。図13は，Yarbus (1967)によって報告された眼球運動の記録であるが，その眼球運動の注視頻度と，上述の本実験におけるそれとが，きわめて類似していることがわかる。一方，Haith, M. M. ら(1977)の発達の研究によれば，生後3～5週間の乳児は，鼻や口をさらに成長した幼児と同程度の頻度で見るとが，目はあまり見えないで輪郭(この場合は，髪の毛の線，あご，ほお，耳)を見るという。それに対して，7週間を過ぎた乳児は，圧倒的に目を見る頻度は高いと報告している。その頻度は，全体の48.9%で，本実験のデータときわめて類似していることがわかる。

3.3 実験Ⅶ：視野制限の条件下における注視頻度と注視順序—その2—

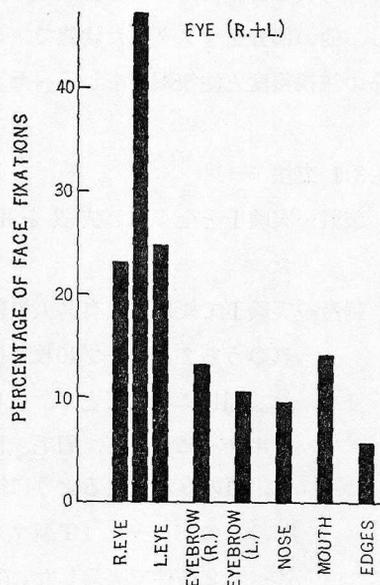


図11 制限視野で顔の部分走査した回数

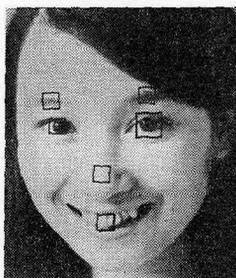
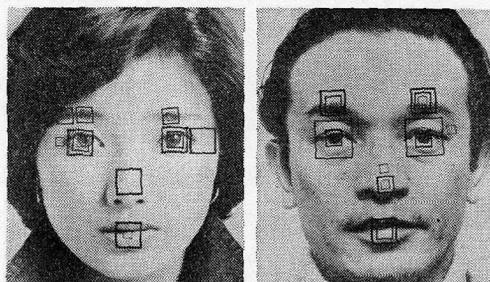


図12 走査回数

- ..... 1回
- ..... 2回
- ..... 5回
- ..... 10回

図12 制限視野実験における顔の各部の走査回数の例。最小の正方形は1回，次の大きさの正方形は2回，3番目の正方形は5回，最大の正方形は10回の走査を示す。

顔写真の目、口、鼻の部分だけを残し、他の部分をマスクした状態で、これらの注視頻度と注視順序をしらべた。

### 3.3.1 方法

装置…実験Ⅰと全く同じ装置を用いた。

刺激…実験Ⅰで使用した有名人の顔写真のうち2グループ40枚を使った。図14に示すごとく、テレビミキサーによって、眉毛、目、鼻、口のみが見えるように操作し、このパターン（T型マスク）を、以下に述べる2条件で実験した。T型の巾は $1.5^\circ$ であった。

手続…条件1；T型マスクの顔写真全体が図14のごとく、5sec間提示され、被験者は誰の顔であるかをできるだけ速く答える（名前以外の表現でも、答として認める）よう求められた。

条件2；被験者にT型マスクの顔写真を $1.5^\circ \times 1.5^\circ$ の視野で観察させ、30secの制限時間内に、できるだけ速くそれが誰であるかを答えさせた。このとき、実験者は“目”、“口”、“鼻”をそれぞれ何回走査したかをチェックした。

被験者…男子3名、女子3名の学生（平均年齢23才）で、全員視力は正常であった。

### 3.3.2 結果

条件1についての正答率は、71%（Appendix参照）であった。条件2のそれは、52%であった。視野を制限したため、条件1に比べて20%程度、正答率が低下した。又、平均反応時間は、条件2の場合、12.1secであった。

第2章で述べた実験Ⅰで、視野の広さが $1.5^\circ \times 1.5^\circ$ の場合は、正答率が約75%であった。本実験の条件2の場合、全体を同時に見ると（本実験条件1に対応）、71%の正答率が得られるので、視野を制限したことによって、約25%の低下（ $0.71 \times 0.75 = 0.53 \approx 0.52$ ）が見ら

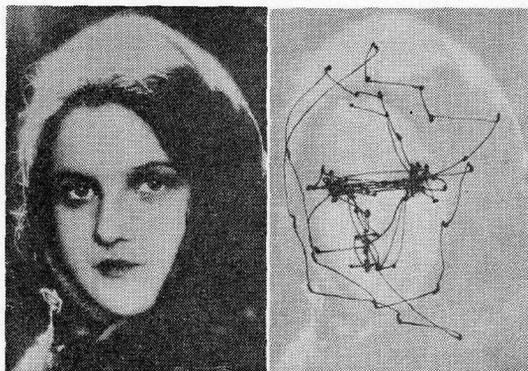


図13 眼球運動の軌跡（Yarbus, 1967より）。

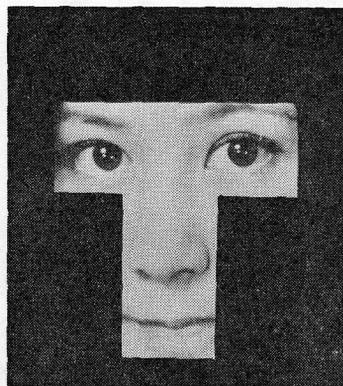


図14 T型マスクの例。巾 $1.5^\circ$ のT型の部分だけ見える。条件2では、このようなパターンを、さらに視野を制限して認識させた。

れる。この事実は、次のようなことを示唆する。すなわち、なんらかの方法によって、刺激パターンを同時に見たにもかかわらず、正答率が  $a\%$  に低下したとする。この刺激パターンをさらに視野を制限して見た場合の正答率は、本論文、実験 I の図 4 の黒四角で示される正答率  $b\%$  に、この  $a\%$  をかけ合わせた値になると考えられる。

一方、一試合中の平均走査回数は次の通りであった。すなわち、左目、右目ともに同数で、それぞれ 4.7 回、口と鼻も同数でそれぞれ 2.6 回であった。口と鼻が同数になったのは、T 型マスクにしたことによると思われる。左右両目を合計すると、9.4 回で、口や鼻の約 3.6 倍よく見ていたことになる。また、注視順序は、実験 VI のそれと同様、左目→右目→口→鼻の傾向が見られた。特に一方の目を見るとほとんどの場合、もう一方の目も見ているということがわかった。

### 3.4 実験 VIII：パターン認知における継時入力の ISI (Inter-Stimulus Interval) および空間的重なりの要因

実験 V において、局所情報が継時的に処理されていることを述べた。また、視野を制限した場合には、明らかに継時的処理を余議なくされる。本実験では、継時的に取り込まれる情報が、いかに処理されて記憶の照合ならびに視覚的知覚が可能なるのかという点についてしらべた。その第一歩として、以下の事項に関して検討を行なった。

(a) パターン知覚における局所情報の刺激間間隔 (ISI) の影響

(b) 継時的に与えられる局所情報の空間的重なりの影響

#### 3.4.1 方法

被験者の知らない顔写真 (大きさ  $5.2^\circ \times 5.2^\circ$ ) 130 枚を用いた。これらの顔写真を、図 15 に示すごとく、横に 2 個、3 個、4 個、5 個、6 個、7 個、8 個に分割して提示した。各画素の提示時間は 100msec であった。

(i) ISI の要因；各画素間に空間的重なりをもたないように分割した。分割数によって、ISI は異なるが、2 個の場合：16, 32, 48, 64, 80, 96, 304, 688, 880, 1104, 2000msec, 3 個の場合：16, 32, 48, 64, 80, 96, 112, 128, 144, 160, 192, 224, 288, 384, 480 msec, 4 個の場合：16, 32, 48, 64, 80, 96, 112, 128, 144, 160, 192, 224, 288, 384, 480msec, 5 個の場合：16, 32, 48, 64, 96, 144, 160, 192, 224, 256, 288 msec, 6 個の場合：16, 32, 48, 64msec, 7 個の場合：16, 32, 48msec, 8 個の場合：16, 32, 48msec であった。

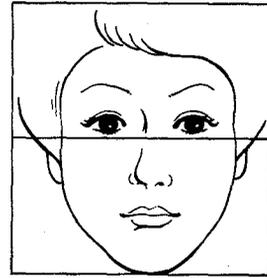
(ii) 空間的重なりの要因；分割数が 2 個の場合についてのみ行った。空間的重なりの大き

さは、全体の10%、20%で、空間的重なりのない場合との比較を試みた。ISIは、96、288、480、688、1104msecの5種類とした。被験者には、継時的に提示された顔が、一つのまとまったものとして見たかどうかを「ハイ」、「イエ」のいずれかで答えさせた。

被験者…男子5名、女子6名が参加したが全員視力は正常であった。

3.4.2 結果

一つのまとまった顔が見えた割合と、ISIとの関係を図16に示した。各点は、全被験者(11名)の平均値である。この図では繁雑になるのをさけるため、主要なISIについての結果を示している。分割数が5個以上の場合、ISIが最も短かいときでさえ、「yes」の割合が50%を越えなかった。50%レベルを、一つのまとまったパターンが見える閾であると定義すると、分割数が2個の場合、ISIが510msec、3個の場合、270msec、4個の場合、110msecに、それぞれ知覚可能閾が認められた。これらの結果は、最初に提示される画素と最後に提示される画素のISIが、常に約500~600msecであれば、全体としてまとまった顔が知覚できることを示している。しかし、分割数が5個以上になると、ISIがいくら短くても、各画素がばらばらに見えるだけであり、まとまった顔が知覚できなかった。たとえば、分割数が5個でISIが16msecのときは、上記の最大時間巾に入るにもかかわらず、ほとんどまとまった顔として知覚されなかった。この結果は、同時にある程度の広さをもった視野の情報が入力される必要があることを示している。空間的な重なりによって、上記の最大時間巾をひろげられるということにはなかったが、分割された画素が少しでも離れていると、もはや一つの顔とは知覚されず、



分割数2の場合

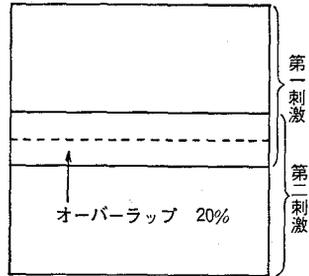


図15 顔写真の分割例。上図は、分割数が2個で空間的重なりがない例である。必ず上から下へ順に各部分が継時提示される。下図は、分割数が2個で、空間的重なりが20%の例である。

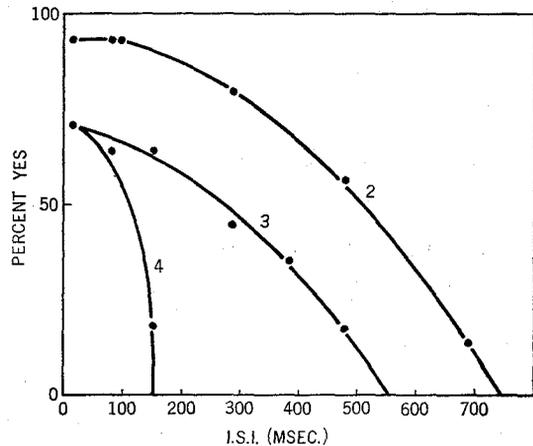


図16 分割画素のISIとパターン全体が一つにまとまって見えた割合。分割数が5個以上の場合、最も短かいISIでも、まとまって見える事はなかった。分割数が2個、3個の場合、最初の画素のISIが500~600msec以内の時は、全体としてまとまったパターンが知覚される事を示している。

上下が別の顔の部分であるような印象を受けた。その意味では、空間的重なりは、パターン知覚において重要な要因であると考えられる。(ただし、本実験の場合は、ディスプレイ上で走査線数本の情報が欠落しているだけであって、顔の長さや大きさは全く変化していない。)一方、重なりが大きくなって、目が2回継時的に提示される場合は、後で提示された目の方が、先に提示された目よりも下に位置しているように見え、あたかも2つの顔がずれて重なっているような、いわゆる複視が生ずる。

### 3.5 考察 (実験V~Ⅷ)

主要な結果をまとめると次のとおりになる。

- (i) 全体的な情報(顔の形、髪型、性別など)処理は、局所的な情報(目、口、鼻の形など)処理よりも速く行われた。
- (ii) 短時間提示においても、目→口→鼻という局所的な情報処理の順序が見られた(Mental Scanningの順序)。これは、眼球運動の解析結果と類似し、視野制限の条件下においても、このような順序が頻繁に見られた。又注視回数順位でも、眼球運動と視野制限条件下とは高い相関が見られた。
- (iii) 視野を制限すると、全体を同時に見た時の正答率に対して、視野の全体に対する比に特定の割合で正答率が低下するようにおもわれる。
- (iv) 部分情報の継時提示において、パターンの知覚可能な時間的、および空間的要因が存在する。すなわち、すべての部分情報は、約500~600msec以内に提示されねばならず、さらに、分割数も4個以下でなければならない事がわかった。
- (v) 分割提示において、部分情報の空間的重なりの影響は見られなかった。

VISは持続時間がきわめて短いbuffer memoryである。この持続時間内に継時提示されたパターンは、時間加重(temporal summation)が生ずる(Jackson & Dick, 1969; Eriksen & Collins, 1968)。これらの実験は、第1刺激と第2刺激が同じ位置に継時的に提示されたときの加重効果をしらべた。VISから特徴が抽出され、形が知覚されて、いわゆる視覚的短期記憶(Visual STM)に貯蔵されるという。Visual STMについての研究は、意外に少ない。特徴抽出されてから、形が知覚されるまでのシステムをVisual Synthesizerとも呼ばれている(Rumelhart, 1977)。明らかに、Visual Synthesizerには、継時処理を含んでおり、この場合、眼球運動ではなく、mental scanningである。実験Vで述べたごとく、顔を短時間提示されたときも、局所情報が継時的に処理されることが認められた。

上述のごとく、実験Ⅷで継時的な統合に関するいくつかの特性についてしらべた。(i)まず、継時的入力(空間的重なり)の要因である。眼球運動によって、注視点を移動させ、全体のパターンを把握する場合における周辺視の機能を考えると継時入力は、空間的に重なった

領域をもっているはずである。そこで、パターン認知過程に、この空間的重なりが必要なのかという疑問が生ずる。この点については、一連の実験成績に示されたごとく継時的入力、少しでも空間的重なりをもっていなければ、一つのまとまった顔として把握されないことが明らかにされた。しかし、重なり量は、パターン認知には、ほとんど影響しなかった。(ii)次に、継時的入力の時間間隔の要因である。実験結果によれば、最初に提示される画素と、最後に提示される画素の ISI が 500~600msec であれば、提示されたパターンは、全体としてまとまりをもって知覚される。ISI がこれよりも長くなると部分がばらばらに知覚される。この結果は、Visual STM の持続時間が約 600msec であるという Phillips (1974)の結果と一致する。(iii)最後に継時入力される部分の空間的広さの要因である。実験Ⅷの結果によれば、分割数が5個以上になると、たとえ ISI が短い場合であっても、まとまって知覚されなかった。このように同時に入力されねばならない広さに制限があるという事実は、実験Ⅰや実験Ⅲの結果と関係しており、興味深い現象である。

#### 4. 一般論議

顔のパターン認識の過程において、局所的な走査が行なわれていることが明らかにされた(実験Ⅴ)。一方、若干不完全であるが、全体的なイメージによって誰かを判断することもできる(実験ⅡおよびⅢ)。以上述べた一連の実験から、全体的な情報処理と局所的な情報処理は、必要とする視野の広さと、処理時間が異なるようである。本研究において得られた成績を考慮すると、顔のパターン認識過程の基礎には、次に述べるごとき情報処理機構の存在を想像させる。

局所的な情報処理系(以下 $\alpha$ 系と呼ぶ)および全体的な情報処理系(以下 $\beta$ 系と呼ぶ)の2つのシステムがある。 $\alpha$ 系は有効受容域が比較的狭く、かつ情報処理速度は遅い。一方、 $\beta$ 系は有効受容域が比較的広く( $\alpha$ 系の約2倍)、かつ情報処理速度は速い。

また、 $\alpha$ 系は各部の詳細な記述は可能であるが、 $\beta$ 系が働かないと、全体のイメージは形成されない。

さらに、 $\alpha$ 系によって走査する場合には、ある程度の順序が決まっていて、それは眼球運動を伴う場合も、Mental scanningの場合も視野制限の条件下でも変わりがなかった。このような走査順序のことを Norton と Stark (1971)は、scanpathと呼んでいるが、少なくとも顔という特殊な対照については、存在していると言えよう。

一方、失認症では、全体的把握と、部分的把握のどちらか一方が、障害を受ける例が多数存在する。<sup>5)</sup>この事実も、上記の $\alpha$ 系と $\beta$ 系の存在を考えると理解しやすい。 $\beta$ 系によって、短時間で全体の概略をつかむことができる。その後、 $\alpha$ 系によって各部分を走査して、詳細な

記述が可能となる。我々のデータによると、各部分がある一定時間内(500~600msec)に、しかも、ある広さの情報が、同時に入力されなければ、全体としてまとまりを失ってしまう。

全体的把握と部分的把握は、異なるシステムによって可能となる。部分情報から全体情報が形成されるのでも、その逆でもない。それぞれ異なる時空間特性をもつチャンネルによって処理されるらしい。

我々の実験および失認症の症例から、mental scannerは、 $\beta$ 型の情報を利用し、部分の全体に対する位置関係を記憶しながら、処理を進めていくのではないかと考えられる。こういう意味で、コロラリー放電(Mackcay, 1972)を再検討すべきであろう。私たちの実験で、複視が生じたり、変形視が生じたりしたのは、Mental Scannerの特性によるのではなからうか。部分の情報が、scannerによって再配列されると考えれば、Parks' Camel(1965)の現象も理解しやすい。mental scannerのより詳しい特性は、現在検討中である。

要約 以上述べた主たる結果を要約する。

(i) 視野の広さが、全体の5%以下になると、パターン認識がきわめて困難であった。視野が全体の18%以上であれば、局所情報の累積によって、誰の顔かがわかった。しかし、全体のイメージは、視野の広さが、33~50%まで広くならないと形成されなかった。

(ii) 全体的な情報は、局所情報よりも速く処理された。又、顔を見る場合においては、局所情報の処理順序が、ある程度中枢で決められているようである。

(iii) 一連の実験から、私たちは、全体的な情報と局所情報とが、異なる時空間特性をもつチャンネルによって処理されていることを示唆した。

あとがき パターン認識は、物体認識、文字認識、風景の認識、顔の認識などに大別される。相貌認知は、古くから全体性が重要であると言われてきた。私たちは、相貌認知を通して、パターン認識のメカニズムの解明を試みてきた。今後ひきつづき、以下の点について、さらに検討をつづけたい。

i) mental scannerのより詳細な特性：特にParksのCamelのように、部分情報が同じ位置に継時的に提示されても、ある条件下では部分情報は、再配列されて全体の形の知覚が生じる。これはおそらく、本文で述べたmental scannerの重要な特性を示していると考えられる。もっとも、この見解に対しては批判的な立場をとる研究者も少なくない(Anstisら, 1967)。

ii) Navon(1977)も指摘するように、部分情報とは、全体の大きさによって規定されるものである。それは、心理物理学における空間周波数チャンネルのように、絶対的に二分されるものではない。本研究においても、 $\alpha$ システムは、全体の情報の約18%を必要とすることが

明らかにされた。このように、認知に必要な部分情報は、全体の大きさによって規定されることは事実である。これは、いかなるメカニズムによるものであろうか。

iii) Moony (1957) によれば、かなり大きなパターンでも、瞬間視でパターンに関するかなりの情報が得られるという。周辺視の広さとパターン認識能力との間には、どのような関係があるのか。

すでに緒言において述べたごとく、私たちは、物体、文字、風景、顔等異なった対象に関連しパターン認識のメカニズムについて、種々の異なった面から着実にデータを集積することによって、人間の情報処理機能の全貌が、次第に明らかにされるものとする。本研究の一部は、本学行動工学講座研究生（昭和52年）米田洋子氏との共同研究である。なお、本論文の要旨は日本心理学会第42回大会（昭和53年10月14日）において発表した。

謝 辞 稿を終るに臨み、本研究の遂行に際して多大な御援助を受けた福島県立医大第2生理学教室の山下良蔵先生、大阪大学基礎工学部生物工学科の末田統先生に厚く感謝の意を表します。また、種々、貴重な御助言をいただいた本学人間科学部行動生理学講座の下河内稔教授、同学部行動工学講座の曾我部正博先生、阪大基礎工学部生物工学科の鈴木良次教授、京都工芸繊維大学心理学教室の秋田宗平教授に感謝の意を表します。また、朝日放送テレビ編成部中畑俊二氏、ならびに本学言語文化部の神田靖子氏には、それぞれ資料および文献の面で御助力いただいたことに感謝いたします。最後に、実験助手として協力して下さった金蘭短大学生の宗保佐和子さん、岸三由貴さんらに感謝します。

（脚注1）視覚的イメージに関する研究は、少なくとも2つの過程について行なわなければならない。一つは知覚—記憶過程であり、もう一つは想像（もしくは再生）の過程である。前者におけるイメージは、*perceptual image* とか、*memory image* と呼ばれ、後者におけるイメージは、*imagination image* とか *mental image* と呼ばれている。実際、現代の認知心理学ではこの両面からイメージの研究が数多くなされている（Pylyshyn, 1973; Kosslyn と Pomerantz, 1977; Anderson, 1978を参照）。ところで一方では、これら2つの過程が同一（もしくは同型）であるということを示す証拠が数多く提出されている（Shepard, 1975; Shepard と Chipman, 1970; Segal と Gordon, 1969; Segal と Fusella, 1970）。

（脚注2）視覚的イメージの持続時間および、イメージ形成に要する時間に関するいくつかの報告をまとめておく。Posner ら(1967, 1969)は、文字照合実験から、刺激の物理的な形の情報を約1.5秒の間記憶していることを報告した。さらに、この記憶は(1)マスキングを受けない点、(2)記憶容量が VIS と比べて小さい点、(3)記憶保持が挿入課題によって大きく影響される点から、VIS と区別し、この記憶を視覚的短期記憶と名づけた。Wilkins

と Stewart (1974) は、Posner らの実験を左右大脳半球における記憶と関連づけて論じている。

一方、Phillips と Baddeley (1971) は、 $5 \times 5$  のマトリクス状のランダムドットパターンを用いて、名前による符号化ができない場合には視覚的イメージは 9 秒間持続すると述べている。また、Phillips (1974) によれば、視覚的イメージの持続時間はパターンの複雑さに依存しているという。たとえば、 $4 \times 4$  のマトリクス状のランダムドットパターンの場合その視覚的イメージは、9 秒間持続するが、 $8 \times 8$  のマトリクス状のランダムドットの場合、3 秒しか持続しなかった。Parks ら (1972) は、挿入課題によって言語化できないようにすると、Posner のパラダイムを用いても、8 秒間持続するという。また、乾と鈴木 (1977) は、心的回転 (mental rotation) の実験においても、文字の視覚的イメージは約 10 秒間持続することを示した。

Posner ら (1969) の実験で、第 1 刺激として音を与えると、聴覚刺激 (音情報) から視覚的イメージ形成までの最小時間の測定が可能となり、その結果は約 750ms であった。

(脚注 3) 宮川洋 (1976) は、情報論的立場から、画像を (1) 記号的画像情報、(2) 情景的画像情報、(3) 情緒的画像情報の 3 つに分類している。(1) は文字であり、(2) は風景、(3) は絵画などを含む。(1) については、情報の流れが画像から受け手へと一方的であるのに対し、(2)、(3) と進むにつれて情報の流れが受け手によって制御され、場合によっては受け手の情報を引き出すアドレスとしての役目しかもっていないと指摘している (宮川, 1976)。人の顔は (3) に近いであろうし、物体は (2) に属すると思われる。最近よく言われる処理の深さ (depth of processing あるいは level of processing) は、(2) や (3) の場合に影響する。 Craik と Lockhart (1972) は、刺激の処理の深さが、物体をどの程度よく記憶するかを決定すると論じた。刺激が入力されると、異なる情報を処理する異なるシステムが継時的、階層的に働くと考えられている。

(脚注 4) これらの事項はちょうど生理学における受容野というものを想定すると考えやすい。ただし、網膜においてその大きさが固定されているのではなく、パターン全体に対する割合で大きさが決まっているという点で通常受容野とは異なる。すなわち、いくつかの部分情報を統合するニューロンの受容野の大きさがだいたい決まっていて、これによって全体の何%かがカバーされている。もし同時刺激される領域が小さい場合にはニューロンは興奮しないが、ある大きさの領域を越えるとその大きさに依存して興奮する。そして、その領域が受容野より大きくなっても応答には変化がみられない。記憶との比較照合において、このような神経システムが働いているということは、さらに押し進めて考えると、私たちが外界のパターンを知覚する際の情報処理様式を示唆しているのではなからうか。あるまとまりをもった部分が同時に処理されていることを示唆する別の現象がある。それは静止網膜像の

消失過程である。網膜に静止した像は消えて見えなくなる。それは決してランダムに消失していくのではなく、また一様に消失していくでもない。必ずまとまりをもった部分がかたまりとなって消えていくことが認められている。(たとえば、乾, 1978参照)

(脚注5) 本実験に関連した諸研究についてみると、たとえば、Wolpert (1924) は、一定の状況図の細部を知覚し得ても、全体の意味は了解できないものを、同時失認 (simultanagnosia) と呼んだ。Luria (1959) によると、すべての失認は、同時失認であるという。Wolpert は、この症候群を部分から全体を形成する機構を含む高次の欠陥に起因するものと解釈した。また、Conrad (1932) によれば、後頭葉損傷による読書障害は、言葉が一つのまとまりをもったものとして把握されず、軽症では、一字一字追っていく読み方になる。一方、頭頂葉損傷による読書障害は、言葉の全体は即座に認識されるが、分節は失われてしまう。語像を即座に、詳細な部分の集積としてではなく、概略 (Umriß) として、把握するのである。V. C. Faust (1947) は、Gestaltzerfall という興味深い症例を報告している。

「……また、あるかなり大きな対象を見たときは、例えばトラックなど、それは即座に、そうしたものと認知することができた。しかし、さらに見つづけると、ひとかたまりの個々の部分、つまり例えば、モーターとか車台、運転室などしか見えなくなってしまう。しかも、個々の部分は、再び一つにまとまりはしないのである (p105, 第2段19行目-25行目) ……まず世界地図をひろげて、ちょうど真中に位置するアフリカを示す。患者は、一瞥するやいなや、即座にこう言う。『これは、アフリカです。でも、すぐ、それは消えてしまい、今はただ色の集りとしか見えません。』患者は何度も目をそらしたりしてから、ちらっと一目見ると、任意にアフリカの全体的印象をつかむことはできるが、ほとんど同時に、それがばらばらになり、一つのカラフルに塗り分けられた国々に分かれてしまうということを示している。休まずじっと見続けると、もはや全体の印象は作られなくなってしまう。(p107, 第2段27行目-40行目) ……タキストスコピックには、語全体を見ているが、普通に観察していれば、すぐその言葉は崩壊する。しかも、すでに読んで理解した時にである。しかしながら、シラブルの部分は正しい構造を残している。それゆえ、経験事象は視覚においては、まず正常人と同じである。それから変化 (Umschlag) が起り、上述の如く後頭葉欠損に特徴的な (das Occipitale) 分節が力を増してきて、個々のものが前面に出てきて、総合 (Gesamtheit), あるいは、全体 (Ganzheit) が再びまとまることはない。(p111, 第2段目58行目-p112, 第1段目5行目)」

#### Appendix

本論文実験1で使用した public face のうち、4グループ80枚の顔写真を約5°の大きさにテレビジョンディスプレイに次の3条件で5秒間呈示し、誰の顔であるかを答えさせた。

条件1 (T型マスク); テレビミキサーを用いて、両側の眉毛、目、および鼻、口の部分

以外をマスクする（本文図14参照）。

条件2（一型マスク）；テレビミキサーにより，両側の眉毛と目以外をマスクする。

条件3；片側の眉毛と目以外をマスクする。以上の3条件についての実験を終えた後，実験に使用した顔写真を80 msの間提示して誰の顔かを判断させ，正答が得られた顔写真に対する結果のみをデータとした。したがって，被験者に顔写真全体を見せた場合には100%誰の顔であるかを正確に判断できることになる。各条件における被験者8名の平均正答率は，条件1で76.7%，条件2で46.5%，条件3で8%であった。

両側の眉毛と目というかなり高い情報をもつ（informative）領域を同時に見ても，正答率が50%に達しないことから，顔の判断に局所的な特徴のみからでは困難であることがわかる。

Goldstein と Mackenberg (1966)は，小学生に同級生の顔を種々の条件下で判断させたところ，小学5年生で，両目だけでは10%，両目と鼻でも47%しか正答が得られなかったと報告している。

---

#### References

- 1) Anderson, J. R. (1978) Arguments concerning representations for mental imagery. *Psychol. Review* 85 pp 249—277
- 2) Anstis, S. M. and Atkinson, J. (1967) Distortions in moving figures viewed through a stationary slit. *Am. J. Psychol.* 80 pp. 572—585
- 3) Blakemore, C. and Campbell, F. W. (1969) On the existence of neurons in the human visual system selective to the orientation and size of retinal images. *J. Physiol.* 203 pp. 237—260
- 4) Brandt, H. F. (1940 a) Ocular patterns and their psychological implications. *Am. J. Psychol.* 53 pp. 260—268
- 5) Brandt, H. F. (1940 b) Ocular patterns as an index of the attentional value of size. *Am. J. Psychol.* 53 pp. 564—574
- 6) Breitmeyer, B. G. and Ganz, L. (1976) Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychol. Review* 83 pp. 1—36
- 7) Conrad, K. (1932) Versuch einer psychologischen Analyse des Parietalsyndroms. *Msschr. Psychiat. Neurol.* 84
- 8) Craik, F. I. M. and Lockhart, R. S. (1972) Levels of processing : A framework for memory research. *J. Verbal Learning and Verbal Behavior* 11 pp. 671—684
- 9) Eriksen, C. W. and Collins, J. F. (1968) Sensory traces versus the psychological moment in the temporal organization of form. *J. Exp. Psychol.* 77 pp. 376—382
- 10) Farley, A. M. (1976) A computer implementation of constructive visual imagery and perception : In R. A. Monty and J. W. Senders (Eds.) *Eye Movements and Psychological Processes* Lawrence Erlbaum Associates, pp. 473—490
- 11) Faust, V. C. (1947) Über Gestaltzerfall als Symptom des parieto-occipitalen Übergangsbereiches bei doppelseitiger Verletzung nach Hirnschuß. *Nervenarzt.* 18 pp. 103—115
- 12) Goldstein, A. G. and Mackenberg, E. (1966) Recognition of human faces from isolated facial features : A developmental study. *Psychonomic Science*, 6 pp. 149—150
- 13) Haber, R. N. and Hershenson, M. (1973) *The Psychology of Visual Perception*. Holt, Rinehart and Winston, Inc.

- 14) Haith, M. M., Bergman, T., and Moore, M. J. (1977) Eye contact and face scanning in early infancy. *Science* 198 pp. 853—855
- 15) Harmon, L. D. (1973) Masking in visual recognition : Effects of two-dimensional filtered noise. *Science* 181 pp. 1194—1197
- 16) Hochberg, J. (1968) In the mind's eye. In R. N. Haber (Ed.), *Contemporary theory and research in visual perception*. New York : Holt, pp. 309—331
- 17) Hoffman, J. E. (1975) Hierarchical stages in the processing of visual information. *Percept. & Psychophysics* 18 pp. 348—354
- 18) 池田光男 (1975) 視覚の心理物理学, 森北出版。
- 19) 乾 敏郎, 鈴木良次 (1977) 視覚における回転図形の認知機構, 大阪大学人間科学部紀要 第3巻 pp. 209—239
- 20) 乾 敏郎 (1978) 中心視における継時処理と同時処理, 大阪大学人間科学部紀要 第4巻 pp. 297—321
- 21) Inui, T., Kawato, M. and Suzuki, R. (1978) The mechanism of mental scanning in foveal vision. *Biological Cybernetics* 30 pp. 147—155
- 22) 乾 敏郎, 宮本健作 (1978) 顔の情報処理様式, 日本心理学会第42回大会論文集 pp. 266—267
- 23) Jackson, R. H. and Dick, A. O. (1969) Visual summation and its relation to processing and memory. *Percept. and Psychophysics*, 6 pp. 13—15
- 24) King-Smith, P. E. and Kulikowski, J. J. (1975) Pattern and flicker detection analysed by subthreshold summation. *J. Physiol.* 249 pp. 519—548
- 25) Kosslyn, S. M. and Pomerantz, J. R. (1977) Imagery, propositions, and form of internal representations. *Cognitive Psychol.* 9 pp. 52—76
- 26) Kulikowski, J. J. and Tolhurst, D. J. (1973) Psychophysical evidence for sustained and transient detectors in human vision. *J. Physiol.* 232 pp. 149—162
- 27) Luria, A. R., Pradina-Vinarskaya, E. N. and Yarbus, A. L. (1963) Disorders of ocular movement in a case of simultanagnosia. *Brain* 86 pp. 219—228
- 28) Mackay, D. M. (1972) Visual stability and voluntary eye movements. In *Handbook of Sensory Physiology*, Vol. VII pp. 307—331 Springer
- 29) Mackworth, N. H. and Morandi, A. J. (1967) The gaze selects informative details within pictures. *Percept. & Psychophysics*. 2 pp. 547—552
- 30) McConkie, G. W. and Rayner, K. (1975) The span of the effective stimulus during a fixation in reading. *Percept. & Psychophysics*. 17 pp. 578—586
- 31) Meyer, G. E. and Maguire, W. M. (1977) Spatial frequency and the mediation of short-term visual storage. *Science* 198 pp. 524—525
- 32) Miller, G. A. (1956) The magical number seven, plus or minus two : Some limits on our capacity for processing information. *Psychol. Review*. 63 pp. 81—97
- 33) 宮川 洋 (1976) 画像情報論, 電子通信学会誌, 59巻11号 pp. 1162—1171
- 34) Mooney, C. M. (1957) Recognition of novel visual configurations with and without eye movements. *J. Exp. Psychol.* 56 pp. 133—138
- 35) Navon, D. (1977) Forest before trees : The precedence of global features in visual perception. *Cognitive Psychol.* 9 pp. 353—383
- 36) Norton, D. and Stark, L. (1971) Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Res.* 11 pp. 929—942
- 37) Neisser, U. (1967) *Cognitive Psychology*. New York : Appleton-Century-Crofts.
- 38) Parks, T. E. (1965) Post-retinal visual storage. *Am. J. Psychol.* 78 pp. 145—147
- 39) Parks, T. E., Kroll, N. E. A., Salzberg, P. M. and Parkinson, S. R. (1972) Persistence of visual memory as indicated by decision time in a matching task. *J. Exp. Psychol.* 92 pp. 437—438
- 40) Phillips, W. A. (1974) On the distinction between sensory storage and short-term visual memory. *Percept. and Psychophysics*. 16 pp. 283—290
- 41) Phillips, W. A. and Baddeley, A. D. (1971) Reaction time and short-term visual memory. *Psychonomic Science* 22 pp. 73—74

- 42) Posner, M. I., Boies, S. J., Eichelman, W. H., and Taylor, R. L. (1969) Retention of name and visual codes of single letters. *J. Exp. Psychol.* 79 pp. 1—16
- 43) Posner, M. I. and Keele, S. W. (1967) Decay of visual information from a single letter. *Science* 158 pp. 137—139
- 44) Pylyshyn, Z. W. (1973) What the mind's eye tells the mind's brain : A critique of mental imagery. *Psychol. Bulletin* 80 pp. 1—24
- 45) Raphael, B. (1976) *The Thinking Computer : Mind inside matter.* W. H. Freeman & Company
- 46) Rayner, K. (1975) The perceptual span and peripheral cues in reading. *Cognitive Psychol.* 7 pp. 65—81
- 47) Reed, S. K. (1973) *Psychological Processes in Pattern Recognition.* Academic Press
- 49) Rumelhart, D. E. (1977) *An Introduction to Human Information Processing,* Wiley
- 49) 齋田真也, 池田光男 (1975) パターン認識と周辺視野 第36回応用物理学会予稿集 1 p. 166
- 50) 齋田真也, 池田光男 (1975) パターン認識と周辺視野 第37回応用物理学会予稿集 1 p. 64
- 51) Segal, S. J. and Fusella, V. (1970) Influence of imaged pictures and sounds on detection of visual and auditory signals. *J. Exp. Psychol.* 83 pp. 458—464
- 52) Segal, S. J. and Gordon, P. E. (1969) The Perky effect revisited ; Paradoxical threshold or signal detection error. *Percept. and Motor Skills* 28 pp. 791—797
- 53) Shepard, R. N. (1975) Form, formation, and transformation of internal representations. In R. Solso (Ed.) *Information processing and cognition : The Loyola Symposium.* Hillsdale, N. J. : Erlbaum
- 54) Shepard, R. N. and Chipman, S. (1970) Second-order isomorphism of internal representations : Shapes of states. *Cognitive Psychol.* 1 pp. 1—17
- 55) Sisson, E. D. (1935) Eye-movements and Schröder stair-figure. *Am. J. Psychol.* 47 pp. 309—311
- 56) Sperling, G. (1963) A model for visual memory tasks. *Human Factors,* 5 pp. 19—31
- 57) Tolhurst, D. J. (1973) Separate channels for the analysis of the shape and the movement of a moving visual stimulus. *J. Physiol.* 231 pp. 385—402
- 58) Tolhurst, D. J. (1975) Reaction times in the detection of gratings by human observers : A probabilistic mechanism. *Vision Res.* 15 pp. 1143—1149
- 59) 鳥居直隆 (1963) 形運動の知覚における眼球運動の測定 日本心理学会第27回大会発表論文集 p. 22
- 60) Von Wright, J. M. (1970) On selection in visual immediate memory. *Acta Psychologica* 33 pp. 280—292
- 61) Watanabe, A. and Yoshida, T. (1973) Roles of central and peripheral vision in pattern perception. *NHK Technical Monogr.* 21 pp. 23—31
- 62) Wilkins, A., and Stewart, A. (1974) The time course of lateral asymmetries in visual perception of letters. *J. Exp. Psychol.* 102 pp. 905—908
- 63) Wolfpert, I. (1924) Die Simultanagnosie-Störung der Gesamtauffassung. *Z. ges. Neurol. Psychiat.* 93 p. 397
- 64) Yarbus, A. L. (1967) *Eye Movements and Vision.* Plenum Press, New York

## SPATIO-TEMPORAL PROPERTIES OF FACE RECOGNITION

T.Inui and K. Miyamoto

The process of visual recognition has been divided into several stages since W. Wundt's work. The pattern fallen on the retina is firstly registered in a transient sensory memory called VIS (Visual Information Storage). The features are then extracted and organized into a visual image. Recently Inui and coworkers (1978) revealed through the psychological experiment that irregularity of the pattern is detected by the parallel processing (in the preattentive process called by U. Neisser) and then the elements of the pattern are processed in serial. They also proposed the neural network model based on the psychological and physiological knowledge. It is, however, unsolved how the parts processed in serial are organized into the whole image. As the first step to solve this problem, the size of the visual field necessary to recognize the monochrome photograph of a face was investigated under several conditions and the time course of the information processing of the face was also examined.

### □ The size of visual field and accuracy of recognition

The subjects were tested on 120 photographs of well-known public faces. All the pictures were full-face against a black back-ground. These photographs were taken by a TV-camera and adjusted to be  $5.2 \times 5.2$  deg in visual angle on the TV display. However, the stimulus was not seen entirely at any one time and only the part within a square area was visible. Subjects can move this window with the control lever at their hand to scan the display freely. They were told that they would be shown some photographs of well-known men. The window size was varied systematically under the following three conditions ;

- (1) The test stimuli to be identified were original photographs with local information.
- (2) The test stimuli were blurred photographs whose high spatial frequencies were cut and which have no local information.
- (3) The test stimuli were original photographs, but, outside the window, the highly defocused photographs were always presented as a background information.

Main findings were as follows ; (case 1) The interaction between size of visual field and reaction time was significant ( $F(5,592) = 51.2, p < .01$ ). Percent correct decreased when the size of visual field was narrowed to 18% for the whole display ( $\chi^2 = 11.9, p < .01$ ). Percent correct was 50% threshold when the size of visual field was 5%. Incorrect responses significantly increased when the size of visual field was narrowed to 18%, while the rate for unidentification was significantly increased when narrowed to 8%. (case 2) The interaction between size of visual field and reaction time was also significant ( $F(3,123) = 29.8, p < .01$ ). Percent correct decreased significantly when the size

of visual field was narrowed to 33% for the whole display ( $\chi^2=16.02$ ,  $p<.01$ ). (case 3) Global background information improved the accuracy of recognition with the limited visual field, even if it was highly defocused.

These results suggested that ; (1) when the size of visual field is less than 5% for the whole display, recognition was very difficult. (2) when the size of visual field is 5-20%, identification is almost possible but the whole image can not be constructed. (3) when the size of visual field is 33-50%, the whole image can be constructed.

#### ② Temporal course of information processing of faces

The facial photographs were tachistoscopically presented to the subjects. Subjects were told to rate about the three local features and three global features. Three local features were rough sizes of the eye, nose, and mouth. Three global features were a rough profile, a rough hair style, and sex. Answers "unknown" were allowed. From linear interpolation, we calculated the exposure duration at which 50% unknown response were obtained and called this duration 50% threshold. Comparing these thresholds for each item, the sequence of the processing will be revealed.

Main findings were as follows ; (1) 50% threshold of the local features were higher than that of the global features. (2) percentage of unknown response for local features were significantly higher than for global features at 16 msec exposure duration ( $p<.01$ ). (3) percentage of unknown response for nose were significantly higher than for mouth and percentage for mouth were significantly higher than for the eye at 32 msec exposure duration ( $p<.01$ ).

These results suggested that ; (1) Global features were processed faster than local features. (2) The sequence of mental scanning, i. e. eye  $\rightarrow$  mouth  $\rightarrow$  nose, would occur frequently in the central processing. This sequence could also be seen in case of scanning with limited visual field.

#### ③ Conclusion

From these spatio-temporal properties of information processing of the faces, we concluded ; (1) Global features and local features are processed through different channels which have different spatio-temporal characteristics. (2) Global features are processed rapidly in the system of large perceptive field and local features are slowly in the system of small perceptive field. (3) The size of perceptive field processing global features is almost twice as large as perceptive field processing local features.