

人間の画像処理過程の動的解析

—情報科学と心理学からのアプローチ

理学部情報科学科助教授 張 善俊

理学部情報科学科教授 後藤智範

外国語学部心理学教授 和氣洋美

外国語学部心理学教授 三星宗雄

1. はじめに

一般に、人が外界から得る情報の9割近くが視覚から得たものであると言われている。見ることによって、ものを認識し、空間的な位置を知ることができる。人間の視覚システムは、今のマシンビジョンシステムより高い性能を持っている。人は視覚から得た情報から、どのようにして外部世界を認識しているのだろうか。人間の画像処理過程の動的解析は、マシンビジョンの自動探査機能の向上や組立ロボットの視覚センサの改良に重要な役割を果たせると考えている。

従来、心理学の立場から定性的に人間の視覚機能について研究を行われてきたが、近年、Marrの視覚理論を起点として、計算論的観点による視覚の研究が盛んになっている。さらに、これらの研究成果に基づいた工学的応用例も報告されている。

本研究は、人間の画像処理過程を定量的に解析することを目的としている。具体的に人間の視覚過程をダイナミックに計測し、情報科学と心理学のアプローチを融合し、眼球運動とパターン認識の関係を解析する。尚、高齢者の加齢による視覚機能の低下を補うために、視線を利用したヒューマンインターフェースの実装も試みた。

人間の画像処理過程に関する研究は、人間の脳における視覚情報処理の未知の機構を探究する手掛かりとして従来から様々な角度から行なわれているが、本研究は、錯視認識における眼球運動の動的特徴解析、バブルマスクによる性別判断とビジュアルサーチにおける眼球運動の特徴などを分析した。以下、本研究報告では、三つの実験を通して、その結果を報告する。また、福祉工学の応用例として、視線を利用したヒューマンインターフェースの実装も紹介する。本研究の一部結果は学術論文誌および学会で発表されているが、未発表の部分については、さらなる研究を重ねて結果を学会で発表する予定である。

2. 視覚情報の抽出と眼球運動

非均一な網膜構造特徴によって、視覚イメージの処理で第一に強調されるのは、視野の中心部であり、特に中心窩にイメージを結ぶ部分に集中している。また、視覚皮質では、ニューロンの50%以上が、視野の10%にあたる中心部の分析に関係しているといわれている。視覚認知システムは、視野の特定部分に目を向けるために、眼を頻繁に動かすこ

とによって、視覚場面のいろいろな部分を系列的にサンプリングをし、視覚情報の抽出に務めている。動きのあるものを正確に追いかけてたり、近いものや遠くのものを見たりと、眼球にはさまざまな運動がある。これらの眼球運動を次に示す。

1. 注視 厳密に「運動」とはいえないが、重要な種類の眼球活動である。
2. サッケード サッケードとは、眼球をある注視から別の注視へと動かす、非常に素早い回転で、断続的である。
3. 追跡運動 これは、動いているものをとらえるための運動で、トラッキングとも呼ばれる。サッケードと違いその動作は低速で、滑らかである。
4. 輻輳と解散 輻輳とは、両眼で近くと同じ対象を注視することができるようにおのおのの眼を内転させる運動である。解散はその反対で、遠くと同じ対象を注視するために視軸平行になるように戻すことである。
5. 補償的運動 この運動は、頭の回転とは正確に同じ角度だけ逆の方向への回転であり、頭の角度を変えても注視し続ける事ができるようにするための運動である。

本研究の第一の実験ではこれらの運動中の注視、サッケードに着目し、注視間の水平・垂直移動量、サッケードの時間特徴、注視点の重要領域からの距離分布などを定量的に計測し、錯視絵を見る際の眼球運動の動的特徴を調べた。第二の実験は人間が顔認識する際に、特に人の性別を識別する際に、注視点はどのような風に分布しているのか、また、人の性別を判断する際に、何処の情報を重視され、どのように総合分析されているかを調べた。第三の実験では、複雑なシーンの中から特定の目標を探す際に、効率よくビジュアルサーチを行うには、人間の視線がどのように移動しているのかについて検討した。最後に、人間の視線情報を元に、視力の弱い人を支援するために、人が見たいところを自動的に検知し、拡大表示させるヒューマンインタフェースを試みた。

2.1 錯視絵を見ている際の眼球運動

2.1.1 実験の説明

本実験では、ヒル(Hill, W. E.)という人物が描いた、有名な二義錯視絵を研究素材とする。この絵は、一見すると若い貴夫人の横顔に見えるのだが、見方を変えると大きい鼻の老婆の横顔とも見ることができる(図1)。この絵を老婆と見ているときに同時に若い女性と見るのは困難である。従来、この錯視絵は生理学、心理学の観点から多くの定性的研究が行われてきているが、本研究では、人が視覚情報からどのようにして対象を認識しているのかを眼球運動の面から定量的に調べている。

実験を20代から30代までの男性12名の被験者に対して実施した。被験者に実験参加同意書を書かせた後、二つのグループに分けられ、1人につき6試行分の実験を行った。実験は次のような流れで行う。

- 1) キャリブレーション：正確に眼球の位置を計測するために、あらかじめ用意した標準パターンを使って、被験者ごとにアイカメラの調整を行う。
- 2) 画像の呈示：安静させた被験者に、画面の中心にあるクロス表示を見せておいて、そ



図1. 誘導錯視画像A-左、B-右 錯視画像-中央

して2枚の画像を継続して表示させ、被験者に表示された絵を自由に見てもらおう。最初は誘導画像を呈示させ、次に錯視画像を呈示させていた。グループAの被験者に若い女性の誘導画像Aを、グループBの被験者には老婆と見やすくするための誘導画像Bを使った(図)。被験者が絵を見ているときの眼球の動きをアイカメラで計測し、パソコンに記録していた。

3) アンケート調査：誘導画像(1枚目)、錯視画像(2枚目)がそれぞれどのように見えたかのアンケートをおこなう。

4) 呈示時間の変更：2、3を1試行とする。本実験では誘導画像、錯視画像の呈示時間のパラメータを三つのパターンに(表1)変更して、被験者一人に2回ずつ、計6試行を行う。

画像呈示パターン	誘導画像	錯視画像	2枚の合計
パターンA	1秒	3秒	4秒
パターンB	5秒	10秒	15秒
パターンC	10秒	10秒	20秒

Table 1: 呈示時間

2.1.2 実験結果と考察

アイカメラで計測されたデータから次の特徴を求めた。

- 注視間移動量の水平成分の平均(水平平均)

- 注視間移動量の垂直成分の平均(垂直平均)
- 停留回数の平均
- 単位時間あたりの停留回数の平均
- 平均停留時間
- サッケードの平均時間
- 重要領域からの距離率

各特徴量は誘導画像表示時、錯視画像表示時、この2枚の画像を見たときの合計時の平均の三つの分類で求められ、それらの比較を行った。なお、アイトラッキングシステムの計測精度は60Hzであるので、1/60秒より早いサッケードでは検出できない場合がある。注視点の定義は、視覚角度1度以下の範囲内で最低停留時間を0.1秒以上停留したものとする。移動量の単位は、アイトラッキングシステムが出力する座標系をそのまま使っている。今回の条件では視覚1度あたり約8という値になる。図2は以前錯視絵を見たことのある人、とない人の注視点位置と停留時間（円が大きいほど停留時間が長い）を表している。

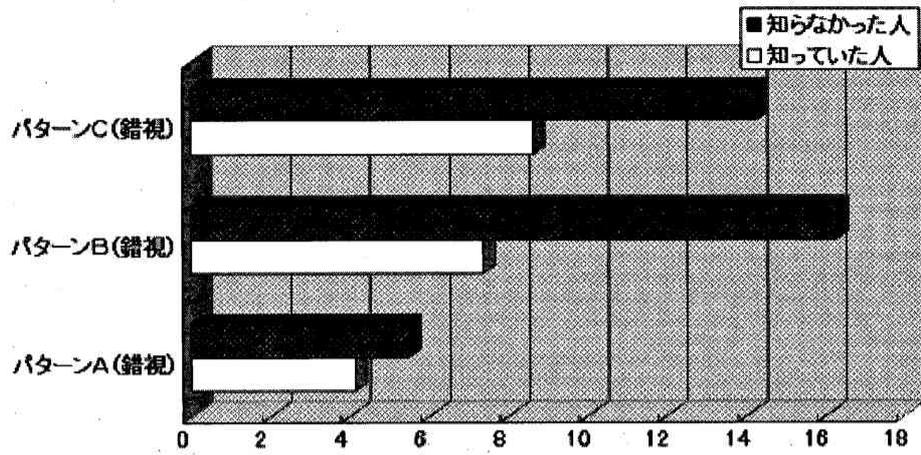


図 2. 被験者の注視点位置と停留時間（左は、以前錯視絵を見たことのある人、右：以前錯視絵を見たことのない人）

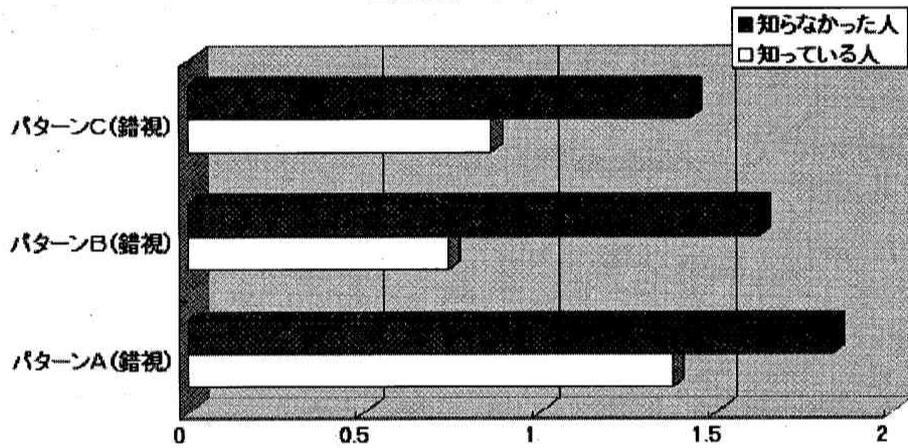
図3、と図4は錯視絵の知っていた人と初めて見る人の眼球運動の動的特徴を示してある。水平移動量では知らなかったグループが表示時間にかかわらず高い値をとっているが、垂直移動量では表示パターンB以外では大きな差はみられない。平均停留時間は表示パターンB、Cでは知っている人の方が停留時間が長いことが分かる。注視点数はやはり画像の表示時間に比例して大きくなっているようであるが、単位時間あたりの注視点数は表示時

間短い方が多く、同条件では知らなかったグループのほうが多い。瞳孔の直径は表示時間によって変わることがあまりないが、やはり同条件では知らなかったグループの方が大きい。サッケードの時間平均はパターンCが大きい値をとっている。

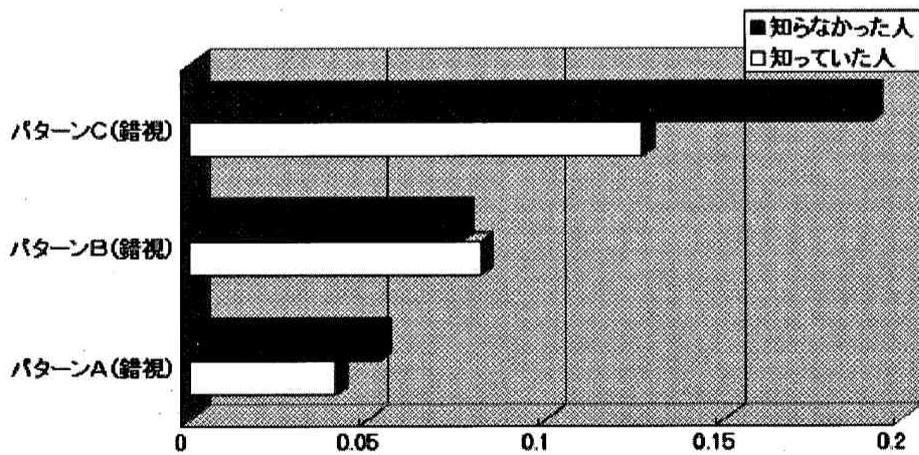
図5は各重要領域にたいする距離率の平均であるが、老婆の目に対する距離率では呈示パターンB以外では知っているグループのほうが高く、老婆の口に対する距離率は呈示パターンA以外では知らなかったグループのほうが高い。若い女性の目に対する距離率は老婆の目と同様呈示パターンB以外では知っているグループの方が高い。偶然の可能性が高いがグラフの形が老婆の目に対する距離率と、若い女性の目に対する距離率がほぼ同じである。錯視画像を見たことがない被験者は、画像の意味を理解するために、割と短い時間に多くの場所を探索している。また、錯視絵が知っていた被験者は画像意味の解釈に重要なポイントに注視点を集中させて、割と長い時間停留させている。なお、見たことのない人が一回目と二回目との見方にも違いが見当たる。それは、画像の解釈に記憶が寄与していることをしめしている。まったく分からない視覚刺激が与えられたときは、サッケードが多く発生させられ、画像の合理的解釈に役に立ちそうな情報の多いところで注視し、一旦合理的な解釈を見つけると、サッケードの数が減り、停留時間が長くなる。



注視点数の平均

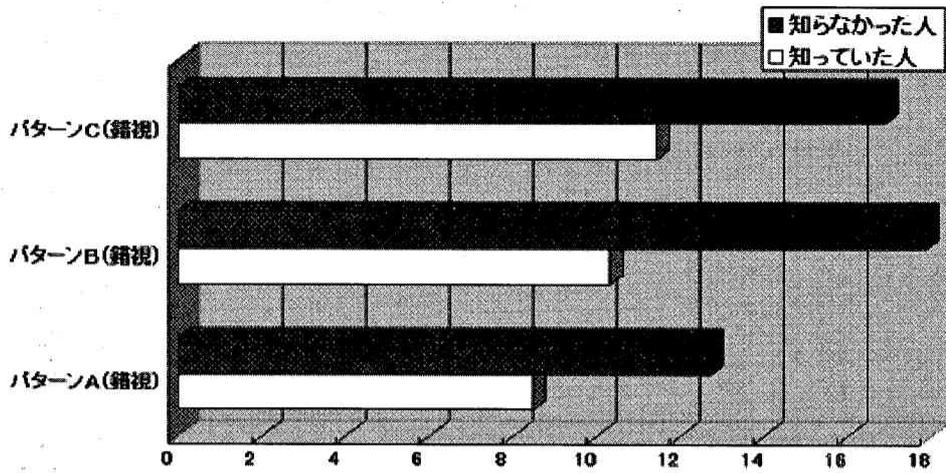


単位時間あたりの平均注視点数

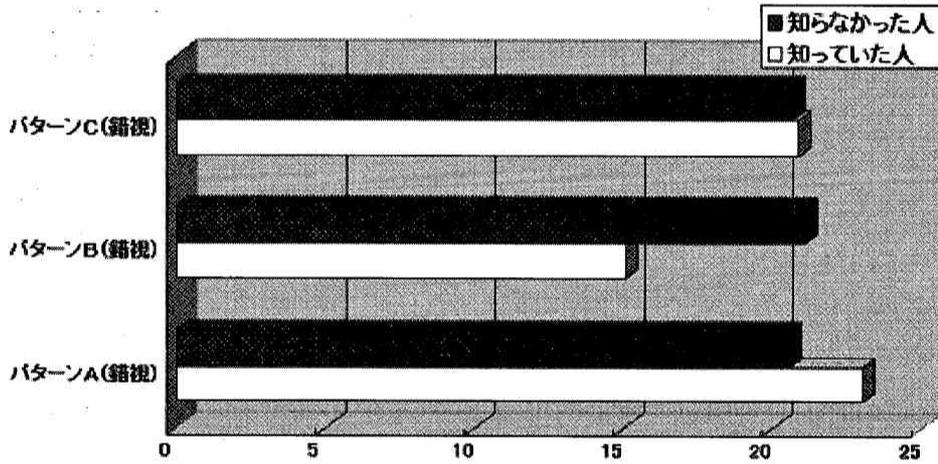


サッケードの時間平均

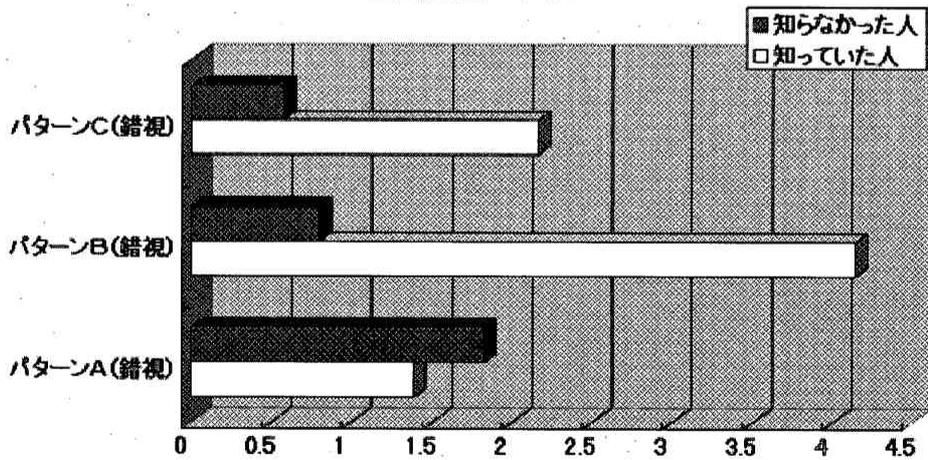
図3、錯視絵を見る際の注視点の統計特徴



水平移動量の平均

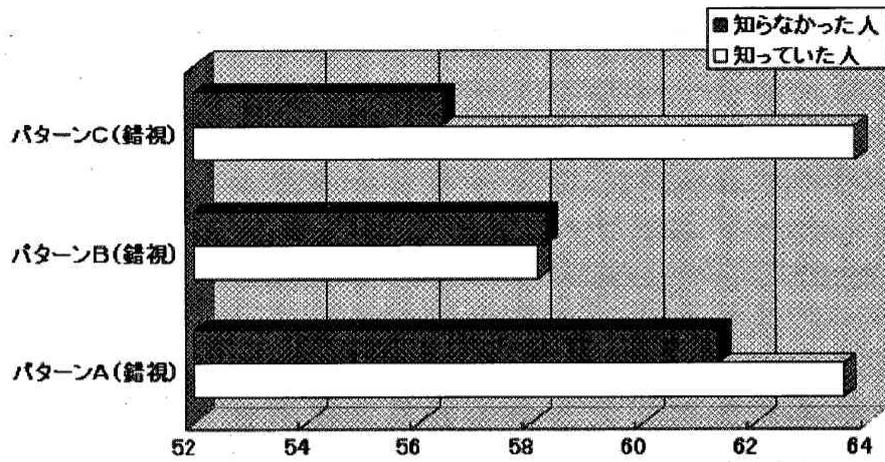


垂直移動量の平均

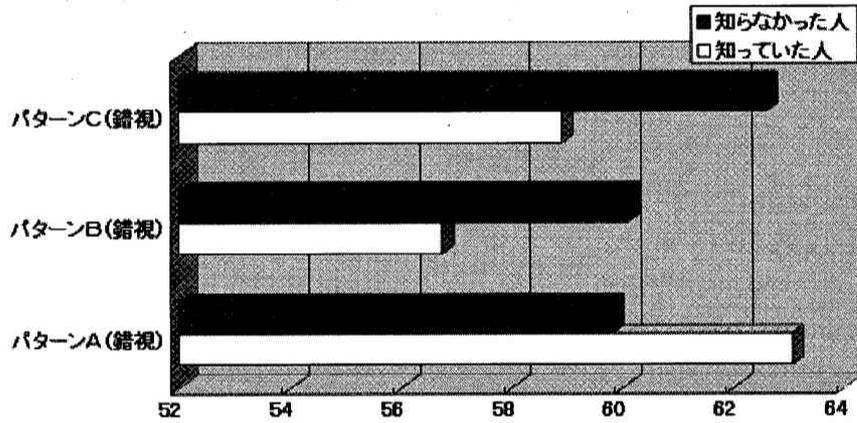


平均停留時間

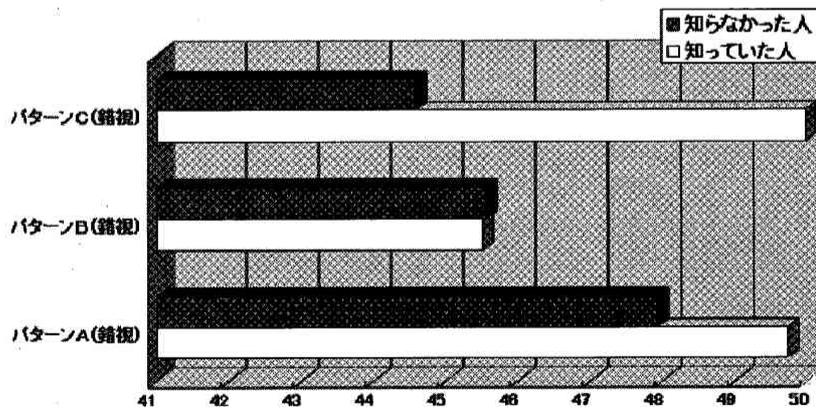
図4、眼球の水平、垂直移動量と平均停留時間の統計特徴



老婆の目に対する距離率



老婆の口に対する距離率



若い女性の目に対する距離率

図5、注視点から重要領域への距離関係

2.2 バブルマスクによる性別判断実験

物体認識において、人は物体を注視することで与えられた画像、状況の中から情報を読み取ろうとする。その時、目的とする情報によって、人の視線移動は異なり、また目的とする情報が含まれている重要な点が異なるために視線の停留点も異なってくる。今回の実験では、特に顔画像データをもとにした性別判断を行うことで、人間が顔を認識する際に視線はどういう風に移動するか、また、人はどの程度の情報量で性別判断が可能かを調べるために行われている。

2.2.1 第一段階の実験

実験は二つの段階で行った。第一段階では、被験者が普通に人の性別を判別する際に視線の振る舞いを調べる。第二段階では、前段階の結果を踏まえてバブルマスクを掛けた画像を被験者に見せて、性別判断をしてもらう。

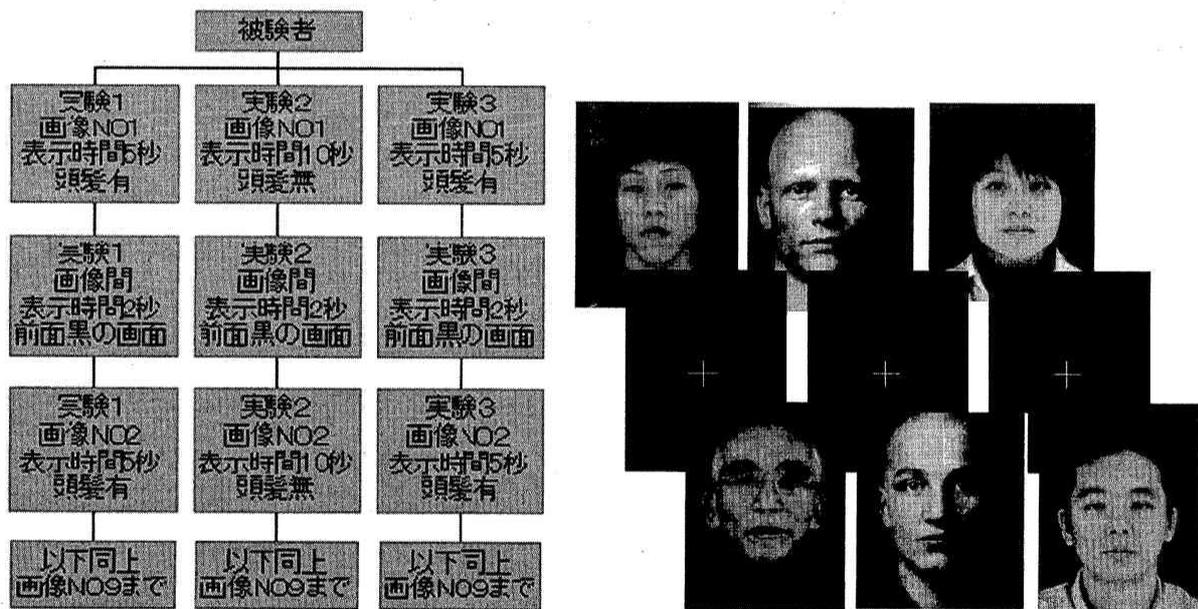


図6、性別判断の第一段階の実験

第一段階の実験の流れは、図6に示されているように、被験者に対して9枚ずつ画像が順番に表示され画像と画像の間に黒地に白の十字を書いた画像が入る。顔画像が移っていない間の時間、被験者はこの白の十字を見てもらう。また、顔画像が表示されている間は画像を見てもらい、黒い画面になったときに指を使って性別を答えてもらう。これは、顎台を使って被験者を固定しているので被験者が答えを口頭で言うことで顔の位置がずれ

ないようにするためである。顔画像は頭髪のある画像は一枚あたり5秒、頭髪のない画像は1枚あたり10秒間表示される。

第一段階の実験結果より、性別判断をする間の人間の注視点分布特徴を調べた。注視点は図7に示されているように、目・鼻・口といったような個所に集中しておりその他の部位の注視点は、以上の点に比べて少ない割合でしかない。(図7)さらに、一枚の顔画像を20の領域に分けて、顔画像が表示されて一番最初に移動した位置を計測したところ、表2のような計測結果を得た。被験者によって多少の個人差はあるものの大きく分類すると、領域の7・13・8・9の四つの領域……つまり目・鼻・口に集中しています。とりわけ領域7は全体の43%、目だけで計算すると全体の63%を占めている。この結果から、眼の部分は性別判断において重要な意味を持っていると考えられる。

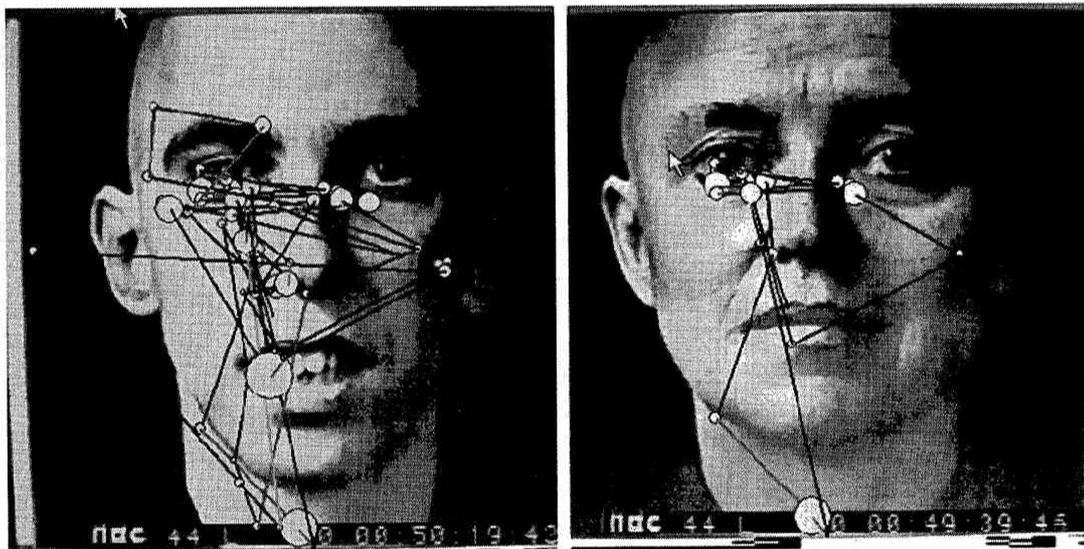
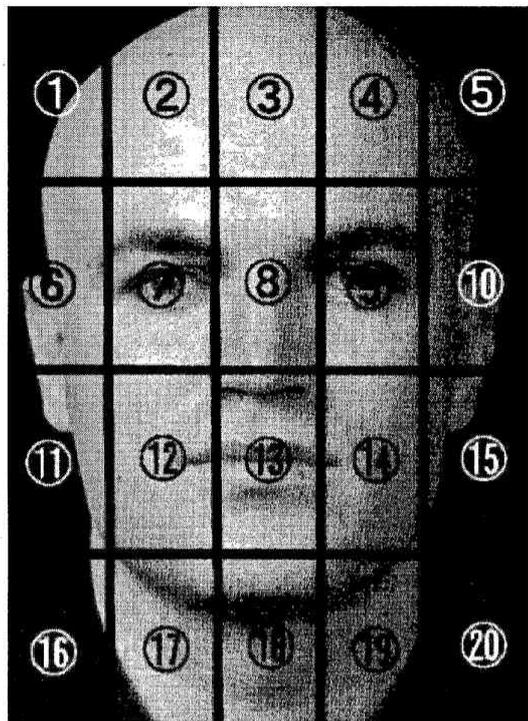


図7、性別識別する間の注視点の分布及び停留時間の分布



	7	13	8	9
被験者1総計	14	0	5	7
被験者2総計	6	8	1	10
被験者3総計	18	5	3	0
被験者4総計	20	2	5	0
被験者5総計	9	4	6	6
被験者6総計	4	12	5	3
全被験者総計	70	31	26	26

	7	13	8	9
実験1(髪有)	23	6	10	11
実験2(髪無)	30	8	3	11
実験3(髪有)	17	17	13	4
総計	70	31	26	26

図8、注視点の分布領域
表2、主な領域の停留点の分布

2.2.2 バブルマスク画像

バブルマスクとは円形の窓のように部分的に見えるマスクのことで、ガウス関数を用いて中心から円形にぼかすように作られている。元画像に5個のバブルマスクを掛けると、図9のようなマスク画像がえられる。第二段階の実験では、マスク画像を利用して限定された部分情報より、人間の性別判断にどのような影響が出てくるかを調べることにしている。

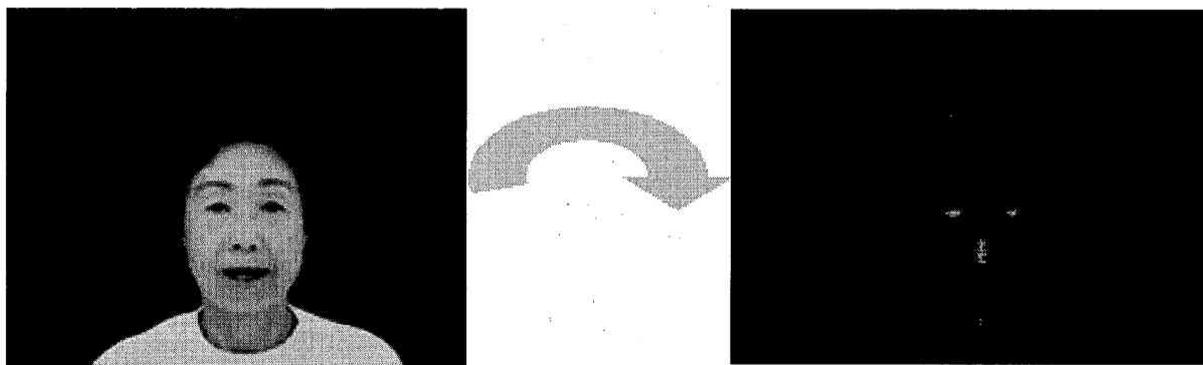


図9、元画像とバブルマスク画像

次の4枚の画像は、ガウス関数のパラメータであるsigmaの値を変えることによりバブルの大きさを変えたものである。

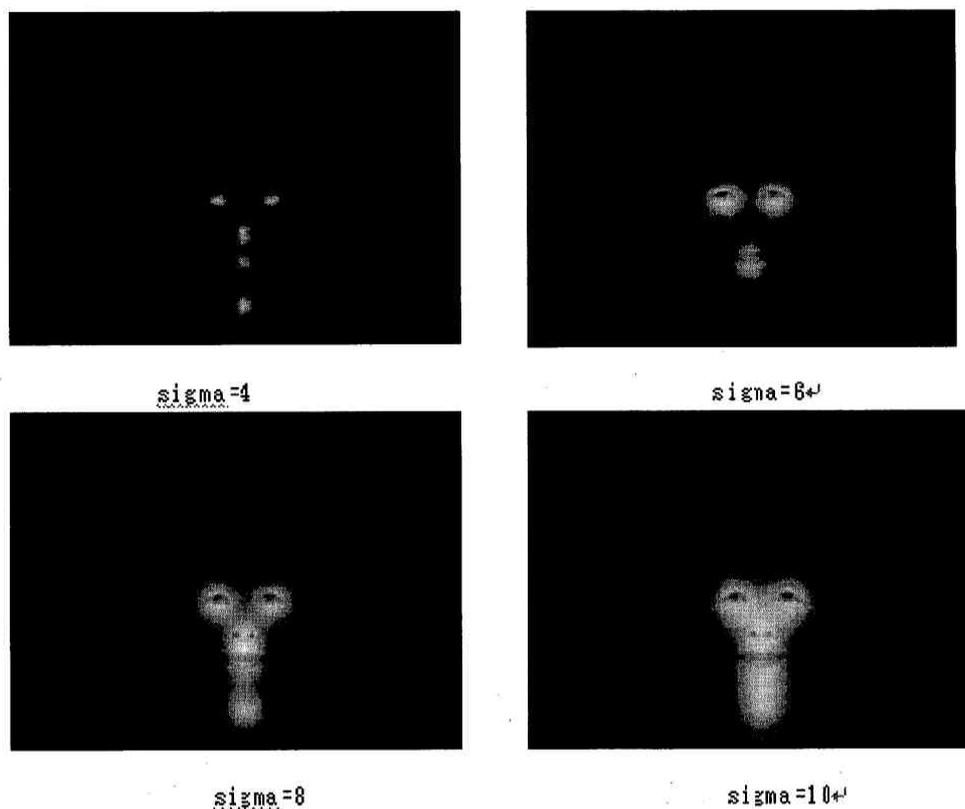


図10. バブルマスク画像の例

第二段階の実験は、このような4段階のレベルのバブルマスク画像を使用している。

2.2.3 第二段階の実験

この段階の実験は実験1から実験9で構成されている。20歳位から60歳くらいと思われる男女300人の顔画像より50人をランダムに抜き出し、その50人に対してバブルの個数別にsigmaの値による4段階のレベルの画像を用意する。さらに、その画像をバブルの個数別にランダムに並べ、その順序で実験を行う。(これにより、完全に同じ画像を無くす。)ただし、バブルの位置は注視点の位置に作られる。そのように準備した画像をノートPCによって表示させる。まず、すべての実験において下図のように前面黒で中央に白の十字の画像が2秒表示され、次にマスク画像が表示される。その後、十字の画像、マスク画像と繰り返される。その十字の画像を表示している間に1つずつ男性ならM(male)、女性ならF(female)と性別を記入してもらう。

実験1～実験3

実験1では下図のようにバブルの個数を5つとし、両目、鼻、口、喉の5点で構成されて

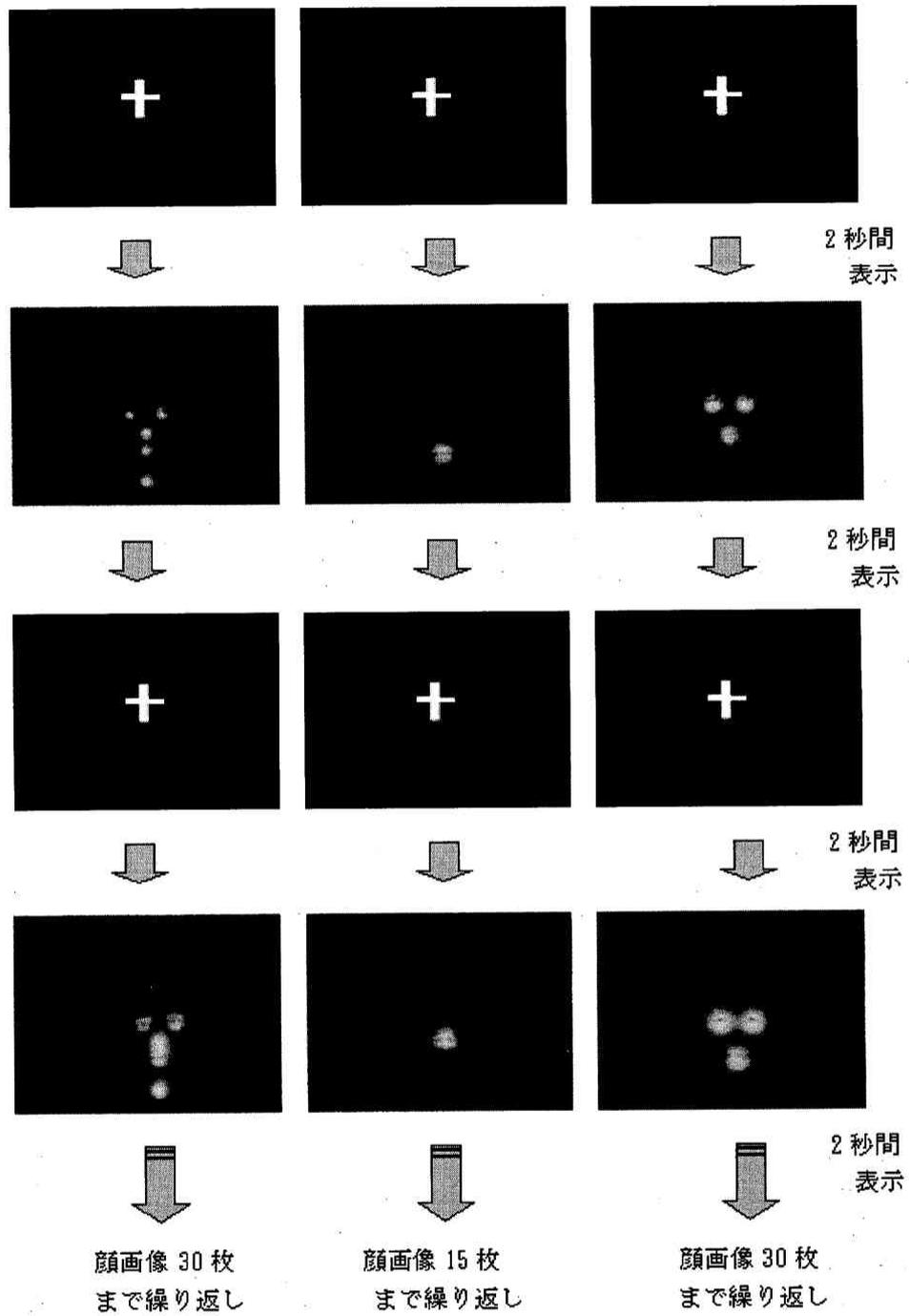


図 12、実験 1～3 の流れ

実験4～実験9

実験1～実験3を1サイクルとする。実験4～実験6はマスク画像の表示時間を1秒に短縮した実験である。実験7～実験9はマスク画像の表示時間を0.5秒に短縮した実験である。以上の9つの実験を43人の被験者に対して行った。被験者は8歳の少年から60歳の男性までの様々な年代の方に協力していただいた。実験の概要は図13に示す。

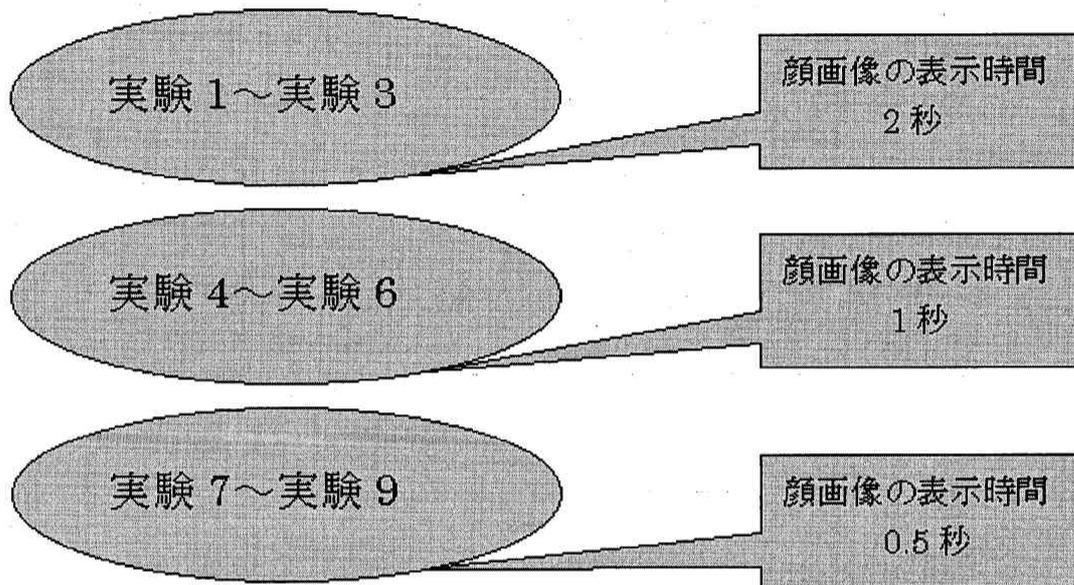


図13、 実験の概要

2.2.4 結果と考察

(1) 画像と被験者別の結果と考察

図14に示す結果より考察すると、実験2, 5, 8の結果より1点での性別判断はなかなか難しいようである。また、実験4, 6の結果より多少の情報を与えられた上での性別判断においては1秒程度の時間で十分判別可能であると考えられる。しかしながら、さすがに0.5秒ではやや難しいようである。補足として年代別にも調べてみたが、本実験では10代、20代の学生が多く全体としてバラツキが多いため公正な評価には至らないが、今回の結果としては30代が一番よく(77%)、60代の方がやや劣る(54%)という結果になった。それ以外では全体としてバラツキが少ない(73%)という結果になった。

実験の結果から、被験者の男女別の比較すると、男性は分かりにくい画像を男性と答えてしまいがちである。被験者の男女別、及び画像の男女別についてそれぞれ調べてみた。

図15. の平均の結果より、やはり女性画像の正解率のほうが悪く、分かりにくい女性画像を男性と答えてしまったためと考えられるので、先ほど述べたように男性と答えてし

まう傾向がややあるのではないかと考えられる。

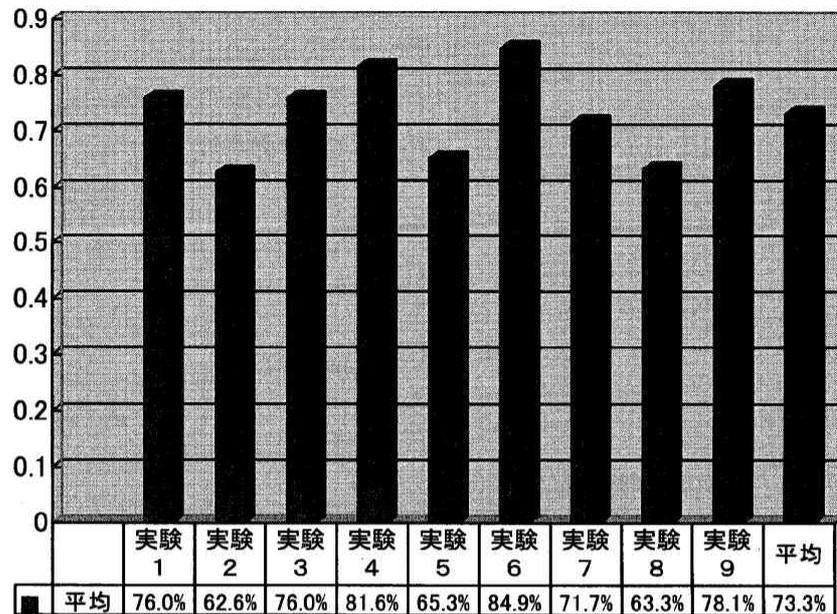


図14、 被験者全体の正解率

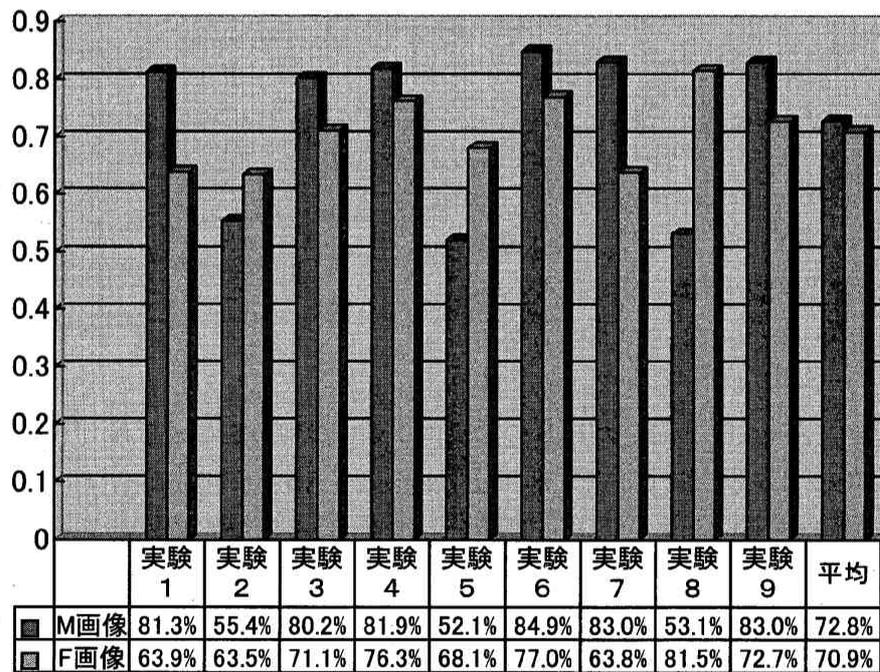


図15、 被験者が男性で画像の男女別正解率

一方、女性の場合も女性画像の方が正解率は悪いが差は少なく、男性のように分かりに

くい画像を男性と答えてしまう傾向があるとは言い難い結果となった。先ほどの結果と合わせると男性画像の方が判別しやすいということがわかった。(図16参照)

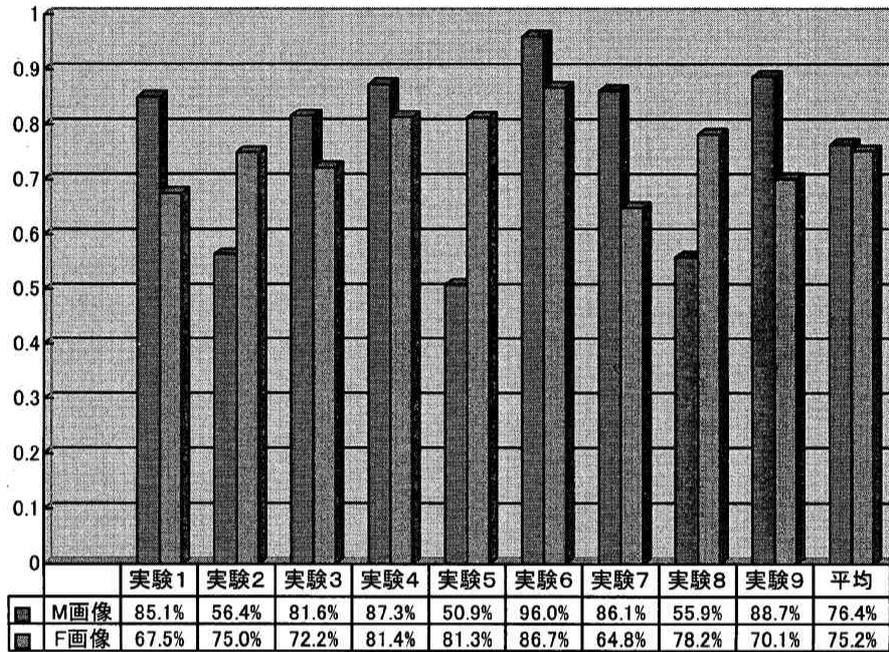


図16. 被験者が女性で画像の男女別の正解率

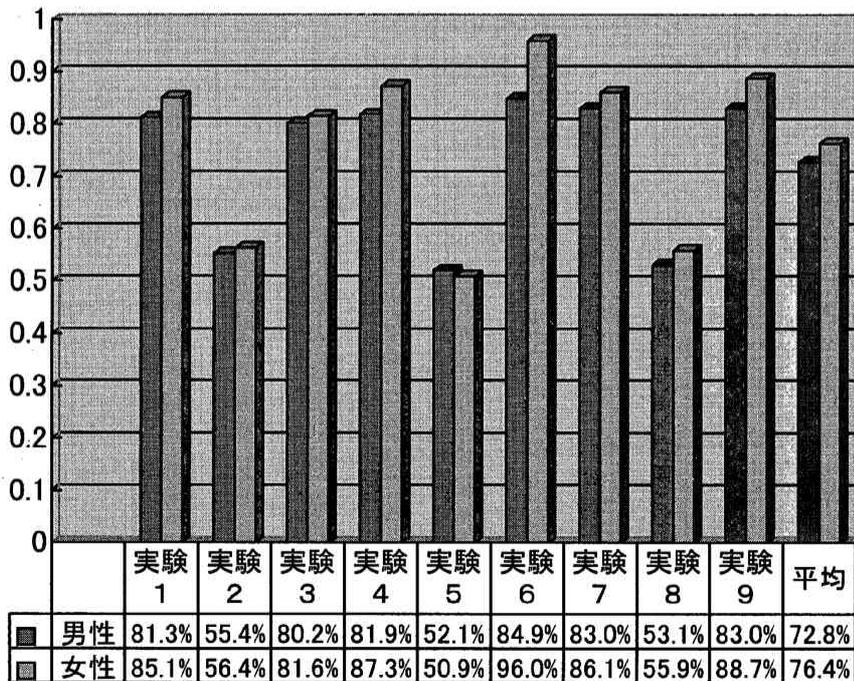


図17. 男性画像における男女別の正解率

さらに、男性画像における男女別の正解率を見ると、図17.を見て分かる通り女性の方がは、正解率が高いことが分かる。女性画像においては、女性の方が正解率がよく、先ほどの結果と合わせると全体として女性の方は、正解率が高いことが分かった。また、この女性画像については男性と女性の差が大きくなっていることについて考えてみたが、推測であるが、普段の生活の中でも女性は女性の化粧などの変化によく気がつくが男性は気が付かないことが多いということに関係があるのではないかと思われた。このことから、女性はちょっとした特徴を捉える能力に優れているのではないかと考えられた。

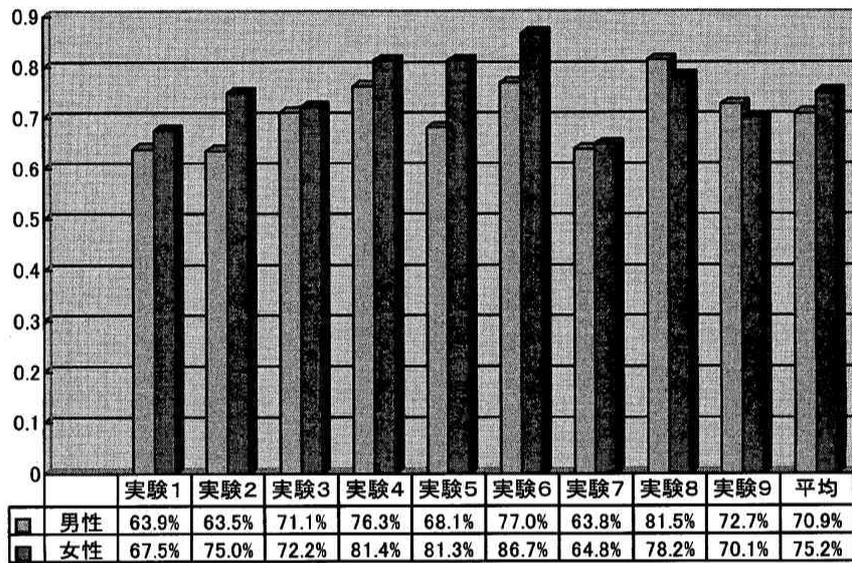
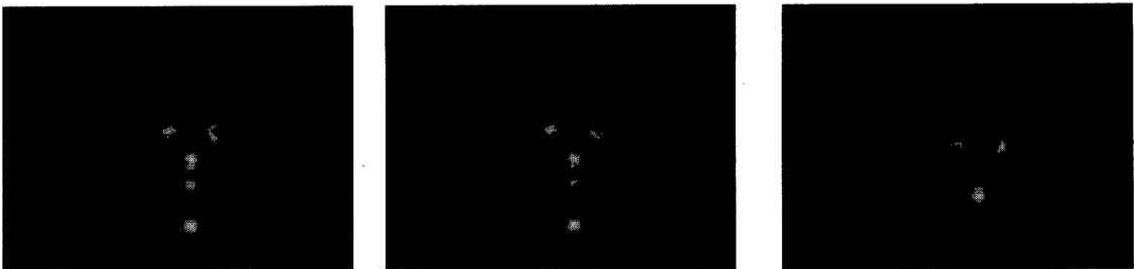


図18. 女性画像における男女別の正解率

以下の画像は正解率が特に悪かったものである。上が本実験で使用したマスク画像であり、下がそれぞれの元画像になる。その元画像を見て分かるように中性的な人が分かりにくいといったわけではなく、元画像ではあからさまに性別判断可能な人であってもマスク処理によって非常に分かりにくくなってしまっている人もあることが分かった。

マスク画像





元画像

図19. 判別しにくい画像の例

(2) バブルマスクの種別について

図20は実験2、5、8のバブルの数が1点での結果を基に、注視点の中でも重要な3点（眼、鼻、口）の内、どこが1番重要であるかを調べてみた結果である。被験者の実験後の感想では、みな口が1番分かりやすいということであったが、驚くべきことに結果としては眼が1番正解率が高いという結果になった。これについて考えてみたが、人が顔を見るときに初見の（1番最初に見る）位置は眼のあたりであるという結果があるので、このことから注視点の中でも眼が1番重要ではないかと考えられる。

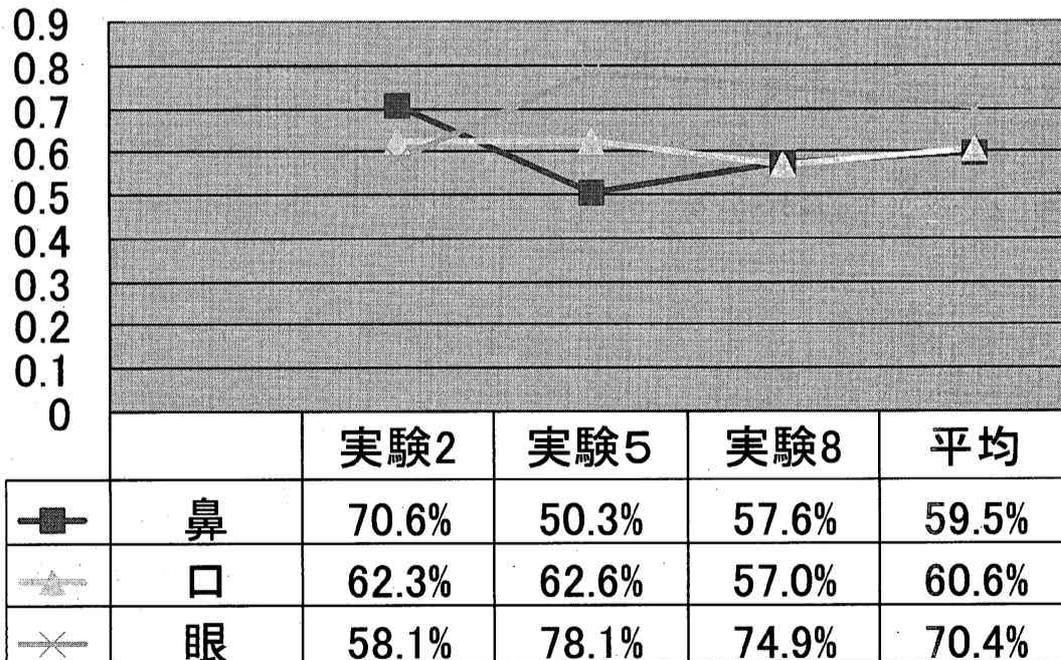


図20. バブルマスクが1個の正解率について

図21はsigmaの値別に正解率を調べ、バブルの大きさの変化でどのような違いがあるのかを調べてみたものである。この結果からsigma=4レベルでは基本的に判別しがたいという結果になった。しかし、5点で1秒以上であればある程度の判別が可能であると考えられる。sigma=6レベルでは3点で1秒以上であれば高い確率で判別可能であるという結果になった。sigma=8レベルでは実験8以外でよい結果となっており、これは1点で1秒以上で判別可能という結果になった。最後にsigma=10レベルでは全体的に1番よい結果となっており、どの実験であってもほぼ判別可能であるという結果になった。ただし、実験2で悪い結果となっているのは実験に用いられた画像の組み合わせが悪くこのような結果となったと考えられる。

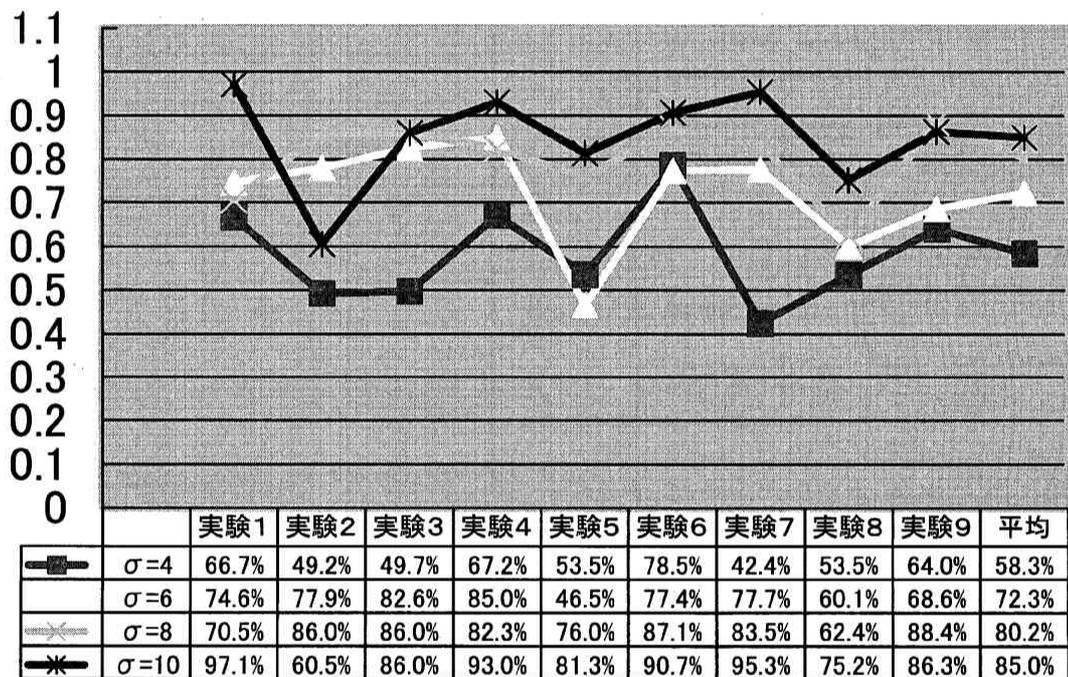


図21. sigmaの値別正解

(3) まとめ

性別判断を行う場合被験者はあの程度の注視点領域から性別情報を得そして判断することが出来る。これは注視点とその付近の有効視野から情報を得て、性別を判断しているのだが極めて短い時間で判断することが可能なことから考えて、被験者が最も効率的に有効視野から性別情報を得ることの出来る注視点の位置が、今回の実験で検出することの出来る注視点の位置になる。つまり、効率良く情報を得るために顔の中心部付近、そして性別などの特徴を示す点として、目・鼻・口に注視点が集まるということである。また、頭髪

が注視点内に含まれていないにもかかわらず、実験結果に大きな影響を及ぼすことから、有効視野内の情報が注視点同様に大きな影響を及ぼすということになる。このことから、性別判断を行う場合は被験者は初見の点とその有効視野内の情報だけで性別を判断することが可能であり、最も効率的に情報を得られる点が目・鼻・口なのである。なお、有効視野については、sigmaの値別の正解率の結果をまとめると、人は多少の情報を与えられた上での性別判断において、注視点である重要な位置（眼、鼻、口など）の情報であるならば

- ◆ sigma=4レベルでは5点で1秒以上でやや判別可能
- ◆ sigma=6レベルでは3点で1秒以上で判別可能
- ◆ sigma=8レベルでは1点で1秒以上で判別可能
- ◆ sigma=10レベルでは全体的にほぼ判別可能

という結果になり、人は非常に少ない情報量から瞬間的に類似する記憶と照らし合わせて判別することができると考えられた。また画像の面積についても調べてみたが、

- sigma=4で5点 —————▶12.3%
- sigma=6で3点 —————▶10.2%
- sigma=8で1点 —————▶4.2%

それぞれこのような結果となり1つのバブルの大きさが大きくなるにつれてバブルの個数が減り面積的にも小さくなっているが、それでも判別できるのはその1点から得られる情報量が大きくなるためであると考えられた。これらの結果よりsigma=8レベルで1点のバブルくらいが有効視野の最低ラインかと考えられる。

本実験では元画像を50人しか用意していなかったため各人物が4, 5回繰り返し現れてしまい、sigmaの値は変えてはいても記憶されてしまうという問題があった。そのため、各実験の結果が公正に評価できるものではなくなってしまった。したがって、すべての実験を通して別の人物を用意して実験する必要がある。また、本実験でも年齢的には幅広く被験者を集めたが10代、20代の学生が多く全体のバランスが取れていなく年代的に評価するに至らなかったため、人数を決め計画的にデータを集める必要がある。

2.3 Visual Searchの基礎実験

我々は、膨大な視覚情報の中からその時点での行動の目的に合った対象を絶えず選択している。このような視覚的選択の中でも、目標とする物体を「探す」ときの注意の特性を解明しようという試みが「視覚探索 (Visual Search)」と呼ばれている。これは、視覚心理分野の研究のひとつの研究分野であり、人工知能や心理学の領域で応用されている。

本実験では、「ウォーリーをさがせ！」という絵本を用いて、提示された画像中から目標を検出するまでの視線運動のデータを解析することによって、探索時の特徴および効率の良い探索法を探る。

2.3.1 実験の内容

図22のように、人ごみや建物などがたくさん描かれている画像の中から「ウォーリー」という人物を発見することで、人間の画像認識と視覚探索過程を探る。その人の視線の振る舞いを視線追跡装置（アイマークレコーダEMR-8 nac社製）を用いて記録して分析する。本実験は、12枚の画像をりようしている。それぞれの画像には、街、海、遊園地など色々な場面が描かれている。



図22、ウォーリーをさがせ

2.3.2 実験方法

実験は神奈川大学生10名の協力で行った。被験者は視線追跡装置を装着し、顎を固定した状態で50cm前方に設置した本からウォーリーを探す。被験者はウォーリーを発見した時点でスイッチを押す。これを12ページ連続して行い、ページ間のインターバルの間、被験者には目をとじてもらった。実験の風景は図23に示してある。

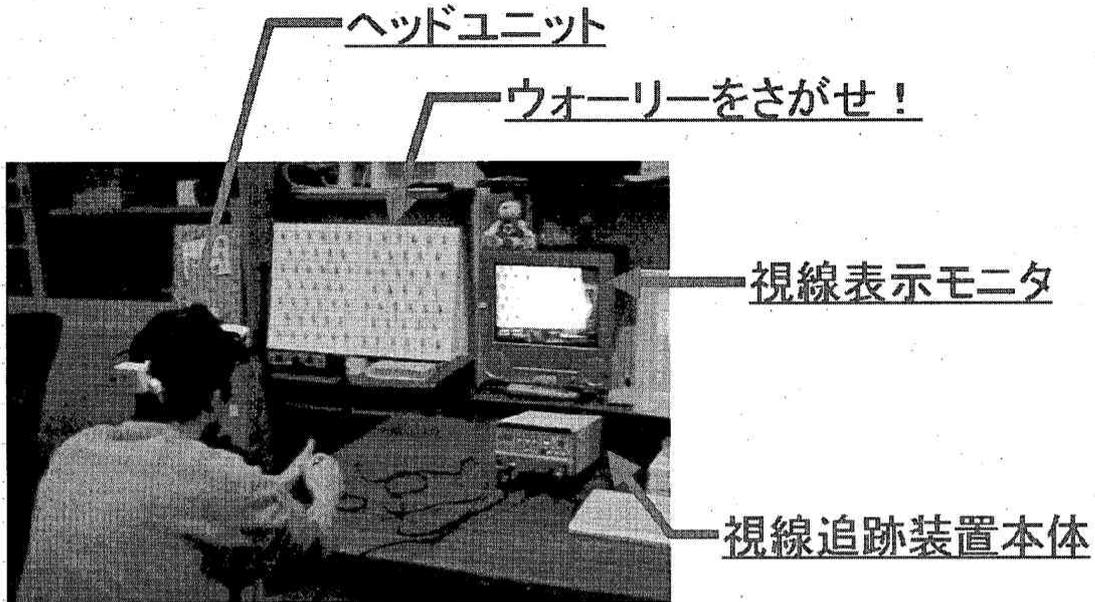


図 2.3. 実験の風景

2. 3. 3 結果と考察

得られたデータから、各被験者がウォーリーを発見するまでの12枚の平均時間をグラフにすると図 2.4. のようになる。このグラフから、データを目標発見が遅いグループと早いグループにわけ、それぞれについて比較する。

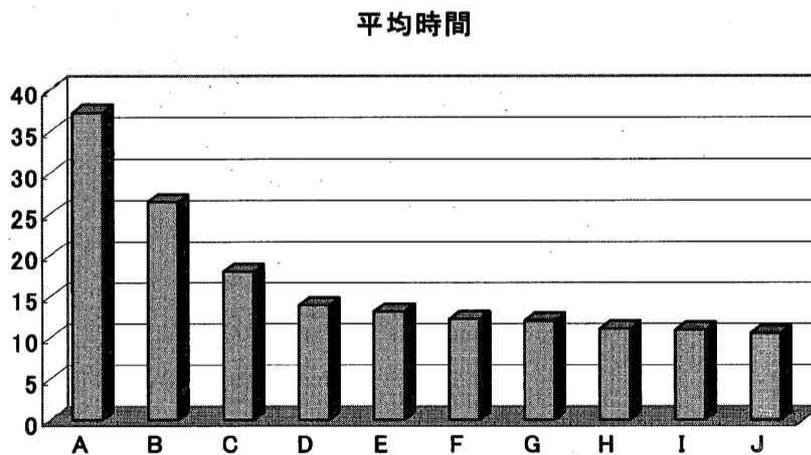


図 2.4. 目標発見の平均時間

まず、提示した画像のうち、比較的建物などの多い画像を探索するときの視線の状態は図25. と図26. のようになった。このように比較していった結果、目標発見の早いグループの被験者は主に人がいる建物のないスペースを見ているのに対して、遅いグループの被験者は全体を探索している、という傾向があることがわかった。

次に、建物が少ない、もしくは人物の多い画像についてみると、この場合の視線軌跡では探索の早さに関係なく全体を見ていて、そのうち約半数の被験者は主に水平方向に探索しているという共通点が見られたので、データを比較してみたが、その場合が特に探索効率が良いわけではなく、効率の良い探索法は見つけることができなかった。(図27.)

次に、経験による探索効率の変化があるのかを調べるため、実験1の被験者のうち4名に実験1と同様の実験を4回行った。それぞれの被験者の平均探索時間をグラフにすると表2のようになる。このグラフから、経験による探索効率の上昇が見られたので、各データの比較、分析を行ったが重要な関係は見つけることができなかった。被験者に感想によると、回をかさねるごとにウォーリーの位置を覚えてしまったという意見が多かったため、強いて言えば被験者が目標の場所を覚えていたことが、探索時間の向上につながったと思われる。

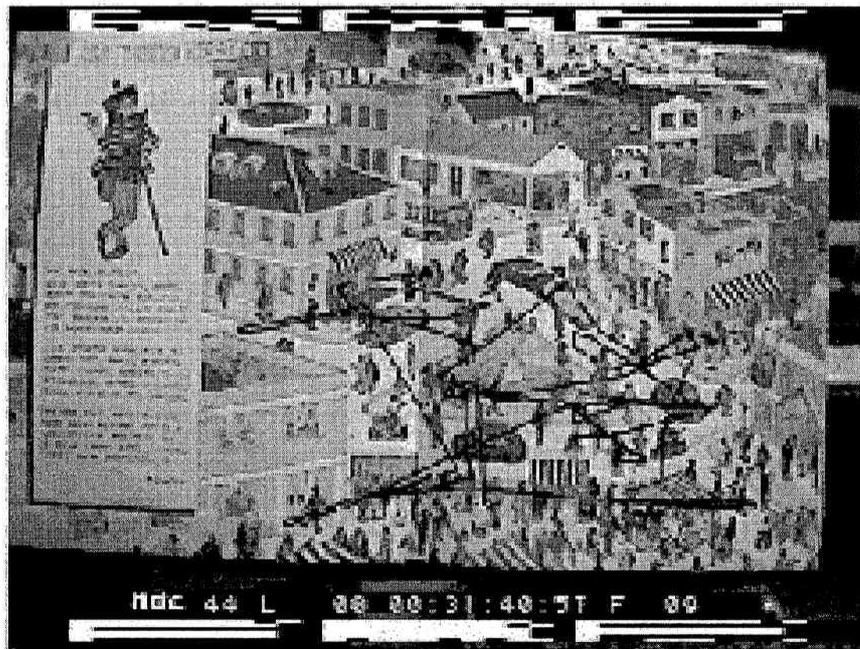


図25. 早いグループの視覚探索経路

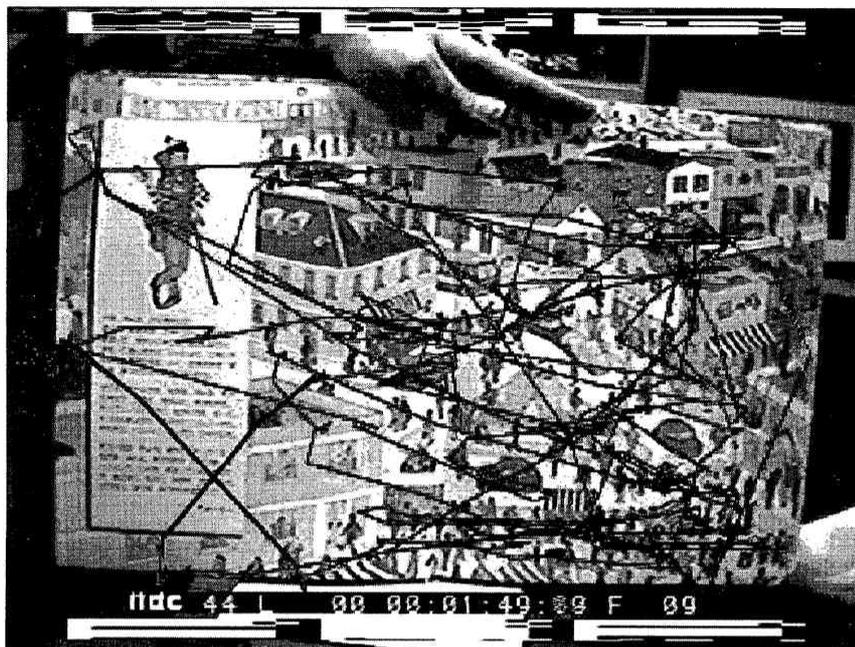


図 26. 遅いグループの視覚探索経路



図 27. 人物が多い画面の視覚探索経路

経験による平均探索時間の変化(秒)

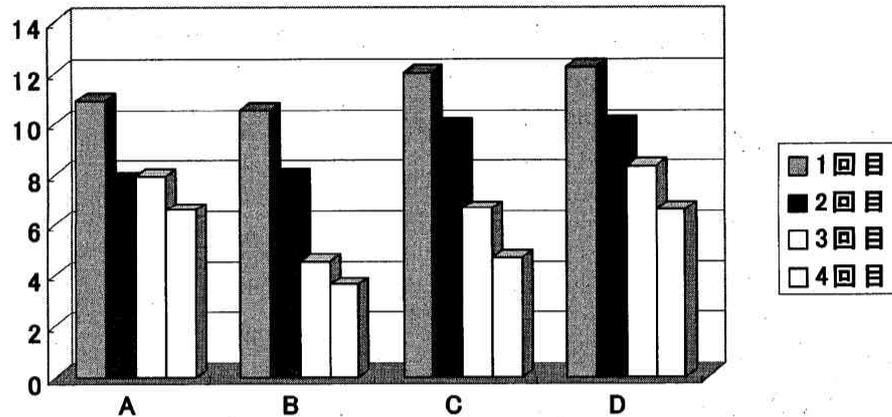


図28. 経験による探索時間の変化

今回の実験からは、建物等明らかに目標探索物と違うものがある場合、その部分を除外して探索すると効率が良いことがわかった。しかし、建物がない場合では有効な探索方法が発見できなかった。

実験2では、経験による探索時間の短縮がみられたものの、実際被験者の記憶力による部分が大きく、法則を発見することはできなかった。実験全体として、データは集まったものの有効な法則の発見には至らなかった。

3. 視線を利用したヒューマンインターフェースの実装

3.1 背景と目的

近年における情報技術の目覚ましい発達には、子供から高齢者まで、幅広い年齢層の人々がコンピューターに接し、利用する機会を増やした。福祉の観点から見ると、視力に問題を抱える方々が、何も違和感を抱くことなく、コンピューターを利用できる環境の整備が必要とされている。そこで、視線を利用したヒューマンインターフェースの開発によって、視力の問題により生じる使い辛さを軽減し、PCの操作を容易することが望まれる。

本研究では、視線の移動を明確に捕らえ、凝視部分周辺の拡大表示を行うシステムの実装を目的としている。

3.2 方法

このシステムにおける処理構成は、“眼球運動の計測”、“計測データの解析”、“凝視部分周辺の拡大表示”の大きく三段階に分かれており、処理の流れは以下ようになる。

一段階目の“眼球運動の計測”では、EMR-8で計測したデータをPCに送信する。二段階目の“計測データの解析”では、受信したデータを解析し、凝視の識別と凝視した座標の特定を行う。三段階目の“凝視部分周辺の拡大表示”では、解析によって特定された座標周辺の画像を取得し、拡大して表示する。

なお、“眼球運動の計測”はハードウェアで、“計測データの解析”および“凝視部分周辺の拡大表示”はソフトウェアで処理を行う。

3.2.1 眼球運動の計測

この段階では、眼球運動を計測し、得られたデータをPCに送信することが目的としている。

計測にはEMR-8（株式会社ナックイメージテクノロジー社製 アイマークレコーダ）を使用した。

アイマークレコーダEMR-8は、瞳孔/角膜反射方式を採用した視線計測システムであり、サブピクセル検出による高精度視線位置計測、装着ズレによる計測誤差を解消し、単眼/両眼の眼球運動計測、瞳孔径の計測等が可能である。

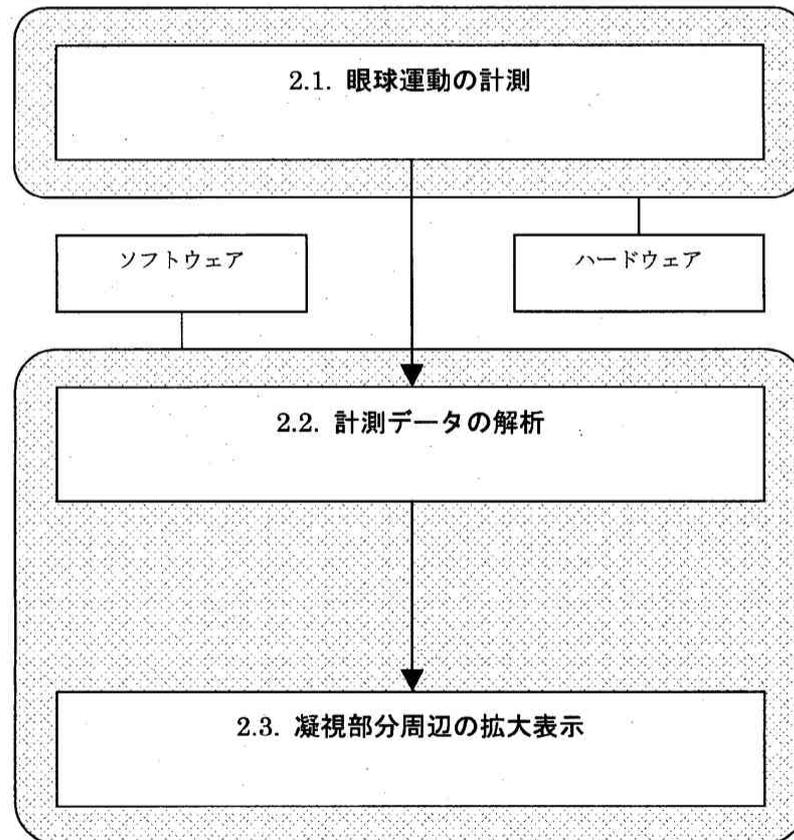


図29. システムの構成

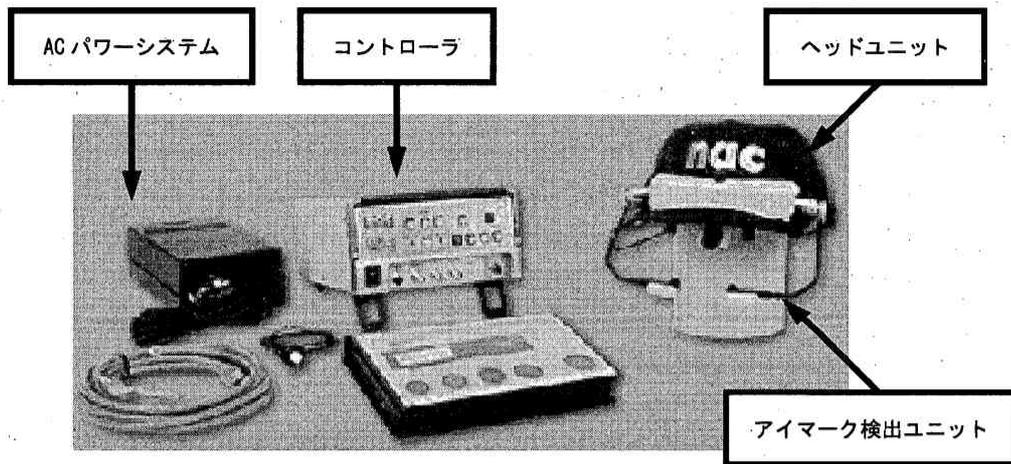


図30. アイマークレコーダ EMR-8

“視線のXY座標”，“瞳孔径”，“フレーム番号”を単眼検出時60Hz、両眼検出時30Hzで検出し、シリアルデータ（RS-232C出力）でPCに送信することができる。

詳細な情報については、アイマークレコーダEMR-8のマニュアル、および株式会社ナックイメージテクノロジーのホームページ（<http://www.eyemark.jp/>）を御覧下さい。

（1）計測準備

EMR-8を使うときは、装着と調整、キャリブレーション、測定という手順になる。普通はヘッドユニットを被る被験者と、セッティングする測定者の二人が必要になる。

1. 装着と調整

まずはヘッドユニットが正しく装着されていることを確認する。

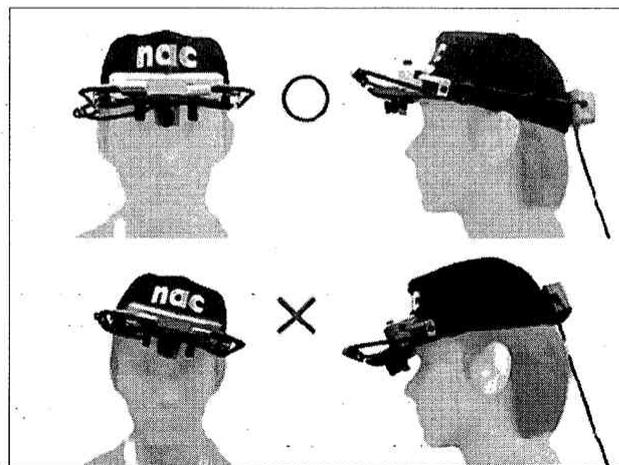


図31. ヘッドユニットの装着

次に被験者の目がモニタの中心に映るようにセンサー位置、ミラー角度、フォーカスを調整する。(図32)

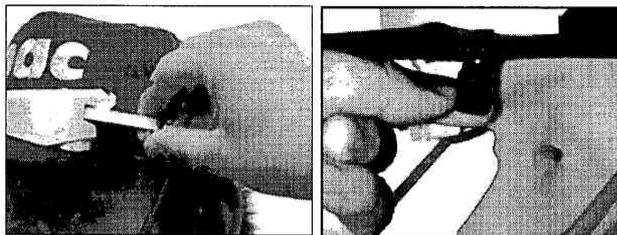
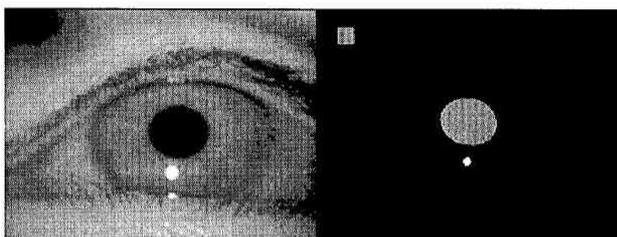


図32. センサー位置調整(左)とフォーカス調整

最後に瞳孔検出映像に切り替えて、角膜反射像と瞳孔が正常に検出されていることを確認する。右の画像は灰色の大きな円が瞳孔を、



白い小さな円が角膜反射像を表している。

図33. 眼の映像(左)と瞳孔検出映像(右)

被験者が眼鏡をしていると、眼鏡のレンズでLEDが反射してセンサーに入ることがあり、測定できない場合がある。(図34)

この場合、ヘッドユニットの被り方や眼鏡のかけ方を工夫することで、撮影範囲からノイズ光を追い出し測定できるようになる。

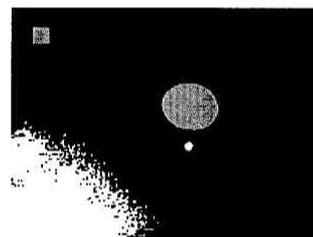


図34. 眼鏡による反射

2. キャリブレーション

被験者の眼の形は各々微妙な違いや、使用の都度変わるヘッドユニットの装着位置は、誤差を生じる原因となり、このままでは正確な計測データを取得することができない。そこで、キャリブレーションによって誤差を吸収し、計測値を補正する必要がある。

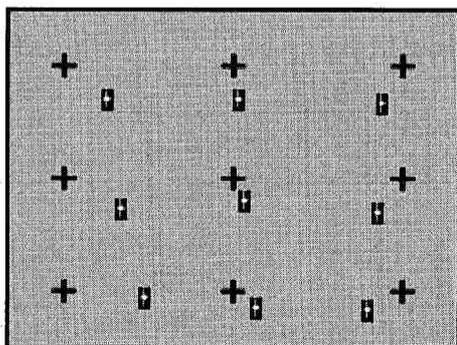


図35. キャリブレーション前

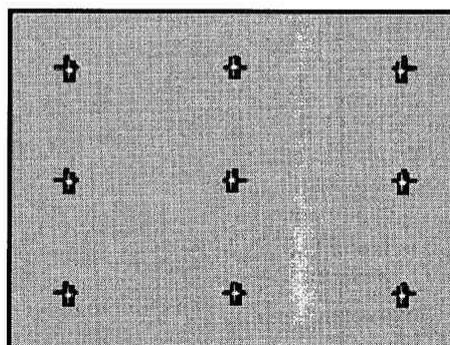


図36. キャリブレーション後

コントローラをキャリブレーションモードにすると視野映像左上に○のターゲットマークが現れる。測定者はモニターを見ながらレーザーポインタで○の位置を指す。(図37)

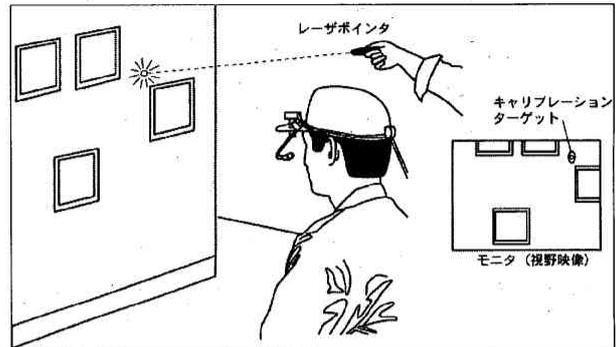


図37. キャリブレーションの様子

被験者はレーザーポインタで指された点を見てスイッチを押す。データを取ることができるとモニター上のターゲットマークが次に移動する。この動作を9点について行う。(図38)

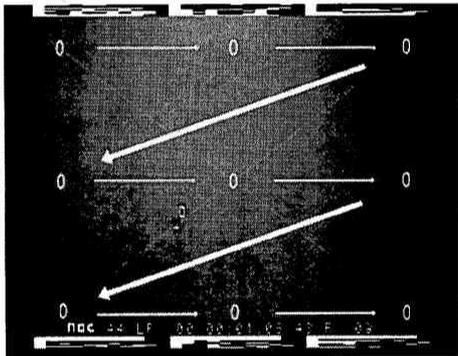


図38. ターゲットマークの移動

これでキャリブレーションは終了だが、確認のため任意の場所をレーザーポインタで指し、被験者に見てもらう。(図39) アイマークが追従していることをモニターで確認できたら測定に進む。

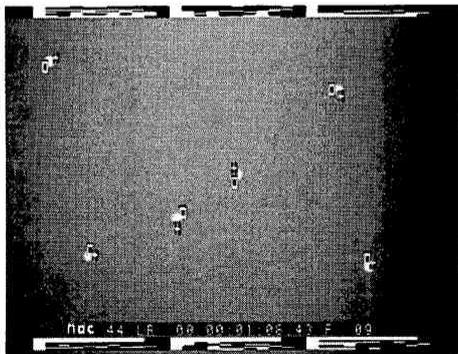


図39. キャリブレーションの確認

3. 測定

被験者は顎を顎台で固定した状態で装着と調整、キャリブレーションを行った後、測定を行う。



図40. キャリブレーションの確認

(2) データフォーマット

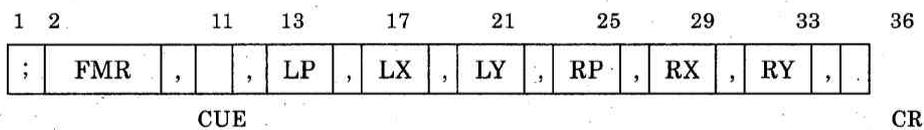
1. 出力データの種類

名称	単位	範囲	エラー時	内容
フィールドカウンタ	1/60 秒	00000000~99999999	---	フィールドカウンタ
CUE	0:OFF, 1:ON	0 or 1	---	キューの ON, OFF
瞳孔径	1/100 mm	000~999	999	瞳孔径 (実寸の 100 倍) 両眼
アイマーク水平座標	pix	000~640	640	画面上のアイマーク位置座標 両眼
アイマーク垂直座標	pix	000~480	480	画面上のアイマーク位置座標 両眼

表3. データフォーマット①

2. 出力データフォーマット

シリアルポートから出力されるデータのフォーマットは以下の通りである。データはASCII文字で出力される。



位置	略称	名称	サイズ
1	;	ヘッダ (セミコロン)	1Byte
2-9	FMR	フィールドカウンタ	8Byte
10	,	データ区切 (カンマ)	1 Byte
11	CUE	キューON/OFF	1 Byte
12	,		1 Byte
13-15	LP	左眼瞳孔径	3 Byte
16	,		1 Byte
17-19	LX	左眼アイマーク水平座標	3 Byte
20	,		1 Byte
21-23	LY	左眼アイマーク垂直座標	3 Byte
24	,		1 Byte
25-27	RP	右眼瞳孔径	3 Byte
28	,		1 Byte
29-31	RX	右眼アイマーク水平座標	3 Byte
32	,		1 Byte
33-35	RY	右眼アイマーク垂直座標	3 Byte
36	[CR]	デミリタ (キャリッジリターン)	1 Byte

表4. データフォーマット②

3. 補足

EMR-8では視線データの検出レートを変えることができないため、プログラム側での変更を可能にした。方法は単純、数回に一度受信データを読み飛ばすことで検出レートを変更する方法を採った。

3.2.2 計測データの解析

この段階では、“眼球運動の計測”で受信したデータから凝視か否かを判断し、凝視であれば凝視部分の座標を求めるのが目的である。そのためには受信したデータである“視線のXY座標”、“瞳孔径”、“フレーム番号”から凝視を定義し、それを元に凝視を識別する必要がある。

“視線のXY座標”は、連続的に捉えることで推測することができそうである。

“瞳孔径”は、凝視との直接の関係は見られないため推測することができそうにない。

“フレーム番号”は、計測開始から何回目に取得したデータなのかを表すだけなので、推測することができそうにない。結果、“視線のXY座標”から凝視を定義しなくてはならない。

(1) 視線座標の選別

右の画像は凝視している状態での視線の移動を表したものである。この画像から、凝視していても視線は上下へ移動しているのがわかる。丸で囲まれた部分は、瞬き後に眼が下を向いた一瞬だけ計測されるもので、推測の妨げとなる。そこで、瞬きの直後に取得したデータは使用しないことにする。

実際に行う処理はというと、EMR-8は瞬きを検出するとエラーとし、PCに対しエラー時の数値を送信する。そこで、PC側でエラー時の数値を確認した場合、瞬きがあったとして直後に受信したデータは使用せず、次に受信するデータを待つ。

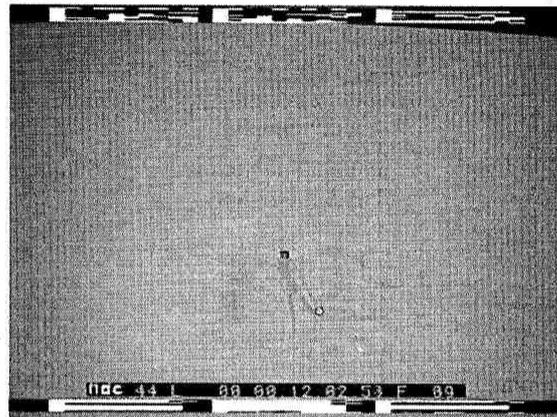
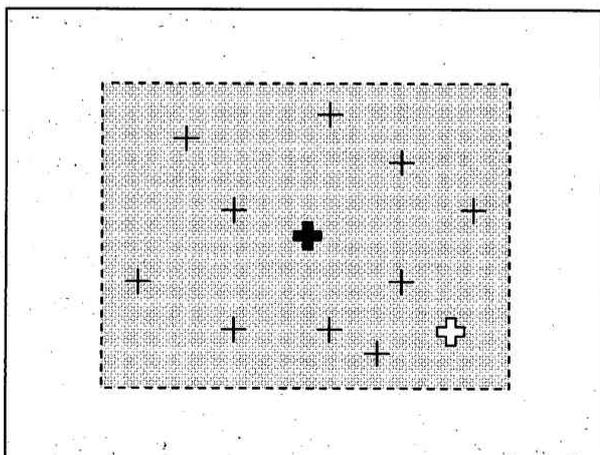


図4.1. 凝視状態での視線の移動

(2) 凝視の定義と識別

実際に計測を行うと、視線は凝視しているにも関わらず眼震によって常に小刻みに動いているのがわかる。物を見ているときの人間の眼球は常に活発に動いており、その動きの中で肉眼ではほとんど捉えることのできない非常に速い周期の揺れのことを眼震と言う。そこで眼震に着目し、眼震による一定範囲内の視線の停留に判断材料とした。



- + ... 以前の座標
- ... 平均座標
- ⊕ ... 直後の座標
- ... スクリーン
- ⋮ ... 眼震許容範囲

図4.2. 凝視の定義

事前の計測で得た一定期間内の座標から平均座標を求め、この座標を中心に眼震許容範囲を設ける。そして、直後に計測された視線の座標が、眼震許容範囲内にある場合は凝視、範囲外にある場合は凝視でないとする。この眼震許容範囲には個人差があり、事前に求めておく必要がある。範囲を狭く設定すると凝視と判断され辛く、広く設定すると凝視と判断され安くなる。

だが、直後に計測された視線の座標を凝視部分の座標とすると座標の変動が激しくなり、結果として拡大表示ウィンドウを表示した際、ウィンドウが激しく動いてしまう。そこで、直後に計測された視線の座標は判断材料とし、途中で求めた平均座標を凝視部分の座標とすることで座標の変動が緩やかにした。常にこの計算を行い、凝視の監視を行っている。

(3) スクリーン座標への変換

EMR-8で取得できる視線の座標は、最大でも640×480であるため、凝視部分の座標がスクリーンの座標に合わないことがある。画面領域のサイズが凝視部分の座標の最大値と同じ640×480であれば問題無いのだが、違った場合は凝視部分の座標をスクリーン座標へ変換し、合わせなくてはならない。

事前に“画面領域のサイズ”を“凝視部分の座標の最大値”で割り、拡大率を求めておく。凝視と判断され凝視部分の座標が得られた場合、拡大率を掛けることで合わせる。例えば画面領域のサイズが1024×768だとすると、得られる凝視部分の座標を1.6倍することで、合わせることができる。

3.2.3 凝視部分周辺の拡大表示

この段階では、計測データの解析によって得られた凝視部分の座標を中心とし、画像の拡大を行うのが目的である。

ここで問題となるのは拡大元画像の取得にかかる時間である。拡大表示に必要なのはスクリーンの画像であり、拡大表示ウィンドウは映りこんではならない。そのため初回の拡大表示以降は、拡大表示ウィンドウの非表示、スクリーン画像の取得、拡大表示ウィンドウの表示という処理を順次繰り返さねばならない。当然この方法では、拡大表示ウィンドウの表示と非表示を繰り返すため、処理に時間がかかるうえにちらつきが発生する。この様に随時画像を取得すると処理に時間がかかり、後の処理に影響が出る。

対処として初回時にのみ拡大元画像を取得し、処理速度を向上させる方法を探したが、結果としてスクリーン画像に更新があっても、拡大表示画像を更新することができなくなった。

実際に行う処理はというと、拡大表示ウィンドウの幅をW、高さをH、拡大率をPとする

と、スクリーン画像のコピーから凝視部分の座標を中心に幅 W/P 、高さ H/P で画像を切り出し、 P 倍してから拡大表示ウィンドウに貼り付ける。

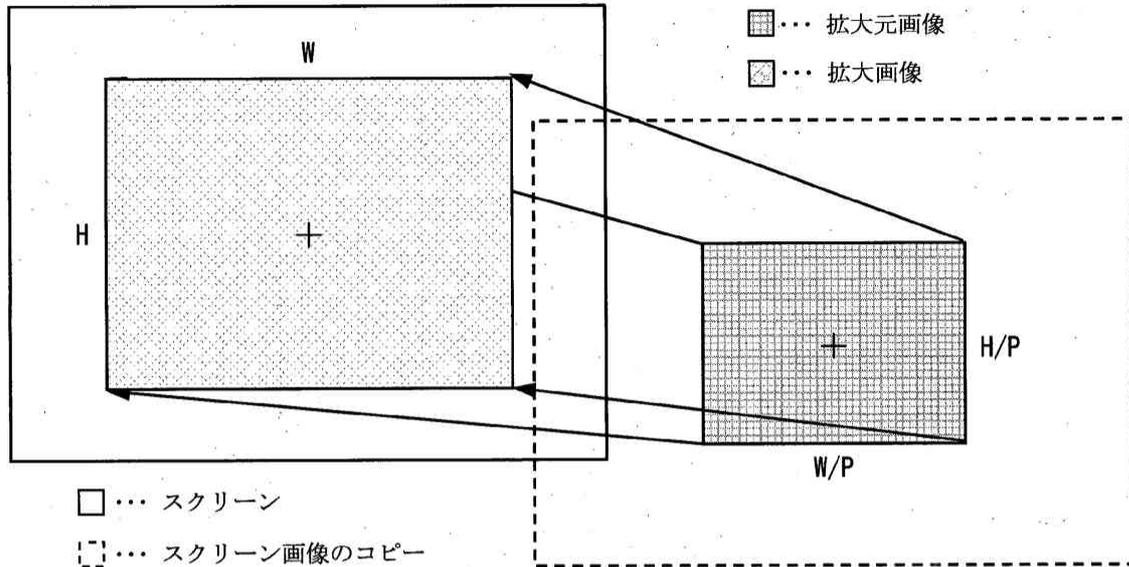


図4.3. 拡大表示処理

現在、拡大率と拡大表示ウィンドウのサイズはプログラム内で変更している。

3.2.4 システムの検証結果

提案するシステムをビジュアルCを使って、ウィンドウズXP上で実現した。システムを検証するために、本研究室の男子学生6名で実験を行った。実験は左眼のみの単眼計測で行う。被験者の顎を顎台に固定し、拡大表示対象PCのスクリーンが、映像表示モニターに映る画像視野映像に収まるよう調節する。その後キャリブレーションを行い、普段のようにPCを操作してもらおう。図4.4はシステム実行の様子を示してある。

3.2.5 考察

本システムは実時間で動作することができた。当初の目標が達成しているけど、次のようないくつかの改善点も残されている。

- ① 開始時は凝視部分を中心に表示されていた拡大画像が、時間が経つにつれて徐々にずれてくる。
- ② 物を見るとき頭を動かすのは自然であるため、顎を固定されると操作しにくい。
- ③ 画面に更新があっても、表示中の拡大画像は更新されない。
- ④ 拡大率を上げると画像の荒さが気になる。

- ⑤ 拡大表示ウィンドウが操作の邪魔になる。
- ⑥ たまに凝視していないのに拡大表示されることがある。

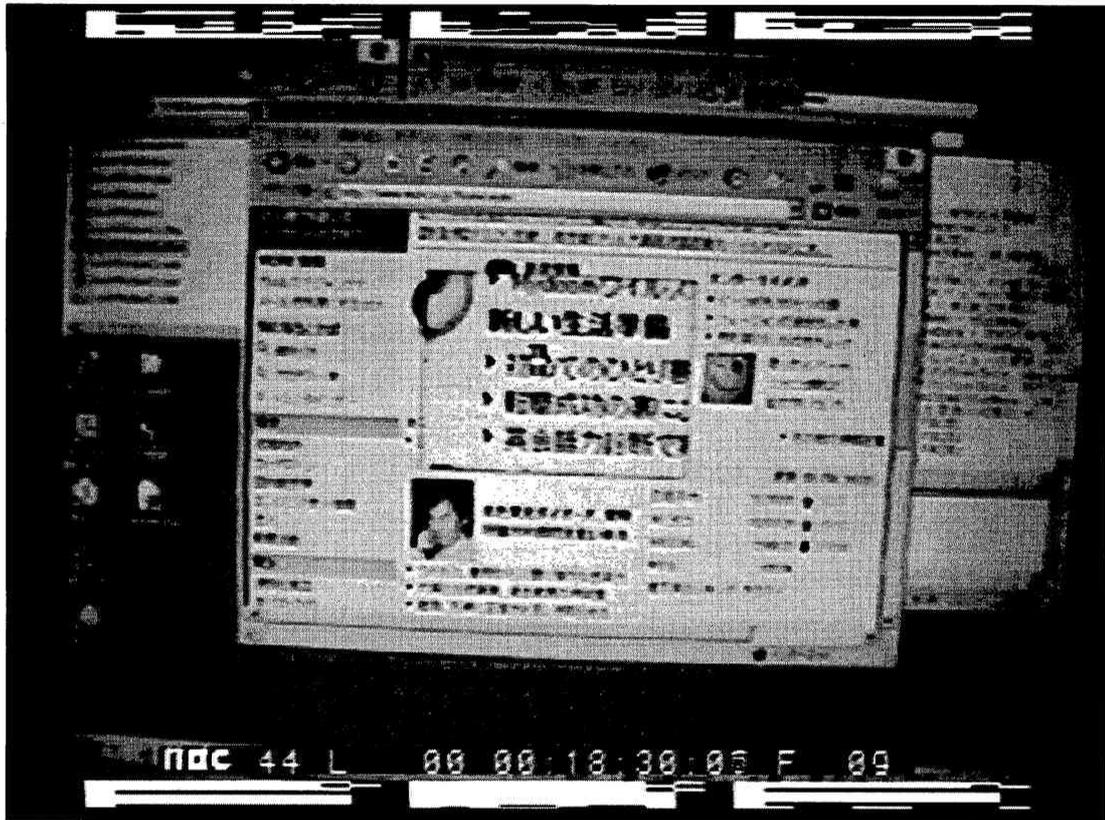


図44. システム実行の様子

①は視線の移動につられて頭が動き、正確な視線の座標を取得できなくなったことが原因である。顎だけでなく頭部全体を固定していまえば、拡大画像が凝視部分からずれることはない。しかし②にもある通り、視線の移動だけで見づらい場合は頭の向きを変えるのが自然であるため、何の固定も無い状態でPCを操作できる方が望ましい。よって、頭部の動きにより生じるズレをプログラムで補正することが課題となる。③の対処として拡大表示画像の随時更新を可能にするには、拡大元画像の取得にかかる時間を短縮する必要がある。単純に拡大元画像取得の高速化が課題となる。④は、拡大表示画像に対して補間を行うことで対処できる。重要なのは受信速度と処理速度のバランスをとることである。⑤は、拡大表示ウィンドウにマウス操作が邪魔されてしまい、PCを操作しづらいというものである。マウスポインタと重なる場合は拡大表示ウィンドウを非表示にするか、拡大表示ウィンドウの上からでもマウス操作を可能にするなど、プログラムの改良が必要になる。

4. まとめ

人間の画像処理過程の動的解析は、様々な学問分野に関連する重要な課題である。それが包括する範囲は、神経細胞の情報処理などのミクロなレベルやブレインサイエンスの基礎、知覚の生理学的・心理学的な側面から、3次元情報処理や人工知能における記憶、推論まで含まれる。本研究は、眼球運動の計測と分析の立場から、錯視認識や、性別識別や視覚探査などにおける視線の振る舞いを考察した。工学的な応用の一つとして、人間が見たいところを発見し、自動的にその部分を拡大して、弱視の人の視覚機能を補う、ヒューマンインターフェースのシステムを試作した。

5. 謝辞

本研究は、神奈川大学共同研究奨励助成「人間の画像処理過程の動的解析—情報科学と心理学からのアプローチ」の助成を得て、2002年4月から2004年3月にかけて2年間実施してまいりました。この場をお借りして、御礼を申し上げます。また、本研究に多大なご支援を頂いた本学関係者に厚く感謝の意を表したい。

【参考文献】

- 1) S. J. ZHANG, K. Y. YOSHINO, "The Dynamic Features of Eye Movement with Illusionary Vision", INFORMATION, vol. 6, no. 5, pp. 593-605. December, 2003.
- 2) 張善俊、吉野和芳, "人間の顔認識時の注視点移動特徴について", IAS 2003, International Symposium of Fusion and Development on Science & Tecnology in the 21st Centry, Tokyo, Japan, pp. 58-64, December, 2003.
- 3) 安田大道、張善俊、前田純治, "アイトラッキングシステムを用いた注視点の移動に関する考察", 情報処理北海道シンポジウム'97講演論文集(97年5月7-8) NO.9, pp. 20-21.
- 4) 田中 昭二, "画像領域の物理的特徴に基づく誘目度評価モデル," 電気情報通信学会論文誌, Vol. J83-A, No. 5 pp. 576-588, 2000
- 5) T. Yagi, K. Gouhara, and Y. Utikawa, "A Computational Study on Fixation and the Effect of an Inhomogeneous Morphological Feature in the Retina", Proc. Of IEEE International Conference on Systems, Man and Cybernetics, Vol. 1, pp. 777-781, 1994.
- 6) V. Bozkov, Z. Bohdaneky, and T. RadilWeiss, "Target Point selection during scanning eye movements," Bio. Cybenetics, 27, pp. 215-220, 1977.
- 7) 青西 享、福島邦彦, "網膜と皮質の不均一性を考慮した注視点移動モデル", 信学技報, NC94-46, pp. 31-38, 1989.
- 8) S. Kemp, "Memoril Psychophysics in Visual Area: The Effect of Retention Interval", Memory and Cognition, 16, pp. 431-436, 1988.
- 9) 大澤幸生, "視覚的インタフェースからのチャンス発見における視点運動," 知能と複雑系, pp. 177-182, 2003.
- 10) J. Tanaka, and J. A. Sengco, "Features and their configuration in face recognition", Memory and Cognition, 25, pp. 583-592, 1997.

- 11) P.G. Schyns, "Diagnostic recognition: task constraints, object information and their interactions", *Cognition*, 67, pp.147-179, 1998.
- 12) F. Gosselin, P.G. Schyns, "Bubbles: a Technique to Reveal the Use of Information in Recognition Tasks", *Vision Research* 41, pp.2261-2271, 2001.
- 13) L. Magnani, G.P. Massara, "Toward a Theory of Visual Abductive Thinking" *Biocybernetics of Vision*, pp.271-289, World Scientific Press, 1997.
- 14) P. Thagard, D.Gochfeld & S.Hardy, "Visual analogical mapping", *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, Hillsdale, Erlbaum, pp.522-527, 1992.
- 15) D.J. Coombs and C.M. Brown, "Cooperative gaze holding in binocular robot vision", *IEEE Control Systems*, pp.24-33, 1991.
- 16) Charles Petzold 著, 株式会社ロングテール 長尾高弘 訳, "プログラミングWindows第5版", 株式会社アスキー
- 17) 北山洋幸 著, "Visual C++ .NETではじめるWin32APIシステムプログラミング", 株式会社カットシステム