映像メディアのための ヒューマンインタラクションの検討

柴田 正啓* 林 正樹* 吉村 俊郎* 柳町 昭夫* 高橋 博#

e-mail:shibata@strl.nhk.or.jp

*ＮＨＫ放送技術研究所
〒157 東京都世田谷区砧1-10-11

#武蔵野美術大学
〒187 東京都小平市小川町1-736

本稿では、人間の自然なふるまいとしての「見る」行為に着目した、映像メディアのための新たなヒューマンインタラクションの試みを紹介する。映像メディアにおける「見る」行為とは、完成した作品を「見る」ことの他に、映像を作るために対象を「見る」という２つの側面がある。この２つに対応し本稿では、実際に存在しない仮想世界を「見て」映像化するための仮想カメラと、見る人の視線によって提示内容が変化していくインタラクティブ番組について述べる。

A study of human-interaction technique for a video medium

Masahiro Shibata* Masaki Hayashi* Toshiro Yoshimura* Akio Yanagimachi*
Hiroshi Takahashi#

*NHK Science and Technical Research Laboratories
1-10-11, Kinuta, Setagaya-ku, Tokyo 157, Japan

#Musashino Art University
1-736, Ogawa-cho, Kodaira-shi, Tokyo 187, Japan

This report introduces our attempts of making new human interfaces for a video medium. *Watching* is a fundamental ability for a human and is a basis for a human-video interaction (HVI). There are two aspects for the watching act in HVI. First, directors watch objects through cameras. Second, people watch video products on screens. According to the two aspects, we have made a *virtual camera* system that visualizes a virtual world and a video display system based on *eye-tracking*.

1　はじめに

テキスト、図形、音声、映像などの表現メディアは、電子化された情報の流通過程において、人間とのインターフェースの部分を担っている。この中でも映像メディアは、近年コンピュータでの動画表示、処理が可能になるとともに、インタラクティブなメディアとしても重要な役割を果たすようになってきた。コンピュータを物理的な媒体とする場合のヒューマンインタラクションでは、キーボードやマウス、GUI(Graphical User Interface)を用いることが多い。この他にもVR(Virtual Reality)と呼ばれる分野のアプリケーションでは、データグローブやHMD(Head Mounted Display)を使って、より多次元できめ細かなインタラクションを実現している。ところが一方で、映像メディアにはコンピュータ以前からの映画、テレビ放送の歴史あるいは流れがあり、この中では「見る人」と「見られる作品」がメディアの構成要素となり、その間にはいわゆる対話性は存在しない。つまり、作品から人への一方的な働きかけのみがあり、人はこれを「見る」だけで、たとえ何らかの感興を覚えたにせよ、それによって作品自体が変化することはない。しかし、「見る」という行為は人間に本来的に備わった能力を使った自然なものであり、人はこの行為自体を意識せずに、作品に集中することができる。このような「見る」という透明性の高い行為に基づく、新たな対話性を映像メディアに持ち込むことはできないであろうか。

本稿では、この「見る」行為に着目した、映像メディアにおける新たなヒューマンインタラクションのための試みを紹介する。映像メディアにおける「見る」行為とは、完成した作品を「見る」ことの他に、映像を作るために対象を「見る」という２つの側面がある。この２つに対応し以下では、実際に存在しない仮想世界を「見て」映像化するための仮想カメラと、見る人の視線によって提示内容が変化していくインタラクティブ番組について述べる。

2　仮想世界を見る　ー　仮想カメラ

映像メディアの歴史は、実世界を記録し伝達することから始まった。カメラの前の世界を撮影し、これを別の場所で再生する。この映像は「実世界の写像」であり、映像で表現される情報は、撮影された実世界そのものの持つ情報の一部である。人間の意図は、何をどのように撮るかといったカメラワークを通して、映像の中に反映される。このような、ごく初期の段階から、カメラは映像だけでなく人間の意図の入力装置としても、映像メディアの中では重要な役割を果たしてきている。

ここで、カメラのヒューマンインターフェースを振り返って見る。現在使われているカメラには、静止画を撮影するスチルカメラ、および動画を撮影するムービーカメラの２種類がある。いずれの場合にも、レンズを通して実際に撮影される画面を確かめながら、カメラポジションやアングルを決めていく。静止画の場合には、シャッターボタンを押すという行為で、ある瞬間のシーンを切り取るのに対し、動画の場合には、被写体の動きに応じてカメラポジションやアングルを連続的に変化させ、時間的な表現も生成する。どちらの場合にもカメラでの撮影行為は、人間が自分の眼で、広い視野を眺めたり、ある物に注目するといった自然な行為とよく対応がとれており、優れたヒューマンインターフェースであるといってよいであろう。

一方で、近年のテレビ番組、映画などの映像制作においては、映像特殊効果が多用されるようになっている。その中でも実写およびCG(Computer Graphics)を使った映像合成は特に重要な技術である。従来、このような映像合成は別々に作成した実写の映像とCGの映像を、特別の装置を使って合成することによって制作していた。この場合に、実写とCGを矛盾無く合成するためには綿密な計算が必要になる。計算通り正確にカメラを動かすためのモーションコントロールカメラ（MCカメラ）が開発され、この作業の精度は向上した。しかし、ヒューマンインターフェースの観点からすれば、人間の見る行為にうまく適合した従来からの

カメラに比べて、同じ映像化を行う道具としては使いにくく、洗練されたものとはいえない。

我々は、制作に関わるユーザーを種々の物理的な拘束から開放し、実写映像とCGが縦横無尽に交錯する世界を映像化する仮想スタジオの実現を目指している。仮想スタジオの中では、実写映像とCG映像を単一の視点から捉え、合成映像をリアルタイムで出力する仮想カメラが重要な役割を果たしている。以下では、仮想世界を「見る」インターフェースを提供する仮想カメラについて述べる。

2．1　MCカメラと映像処理を組み合わせた仮想カメラ

我々はこれまで、実写とCGのリアルタイム合成システムを開発し、実際に映像制作に応用してきた[1]。しかし従来のシステムでは、スタジオ空間の制約などにより、超ロングショットから近接撮影までを連続的にカバーするようなカメラワークは難しかった。これに対し、今回開発したシステムでは、MCカメラと、撮影映像のリアルタイム映像処理装置を、仮想撮影の原理のもとに組み合わせた撮影システム＝仮想カメラを実現している。この仮想カメラでは、実際のカメラを操作するのと同様なインターフェースで、物理的には存在しない世界（仮想世界）を、超広範囲、高速のカメラワークで撮影することができる。

2．2　仮想カメラの原理

MCカメラで撮影した実写の映像とリアルタイムCGをクロマキー合成して出力映像を得る。ユーザーは専用の操作器（写真1）でカメラワークを行い、映像合成された世界を撮影する。この際、MCカメラ自体が物理的に移動できる被写体まわり約3mは、実際のカメラを駆動して撮影し（図1（a））、これ以上遠方にユーザーが指示したカメラ位置が移動したときはカメラ映像を縮小処理することで等価な効果を得る（図1（b））。その結果、被写体の周りのあらゆる位置からのカメラワークを実現することができる。

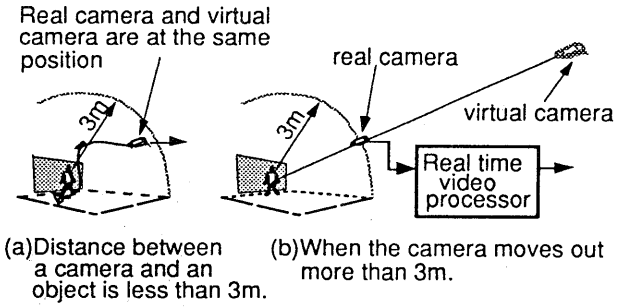


図1　仮想カメラの原理

Figure 1. Principle of the system

図1（b）において必要な映像処理は仮想撮影の原理を使って以下のように導くことができる。いま、ユーザーが指示したカメラ位置における仮想的なカメラである仮想カメラに対し、MCカメラを実カメラと呼ぶと、実カメラ、仮想カメラ、被写体の3者の関係は図2のようになっている。実カメラで撮影した映像を被写体をマッピングしたパネルとして扱うと、仮想撮影とは、このパネルを任意の仮想カメラワークで再撮影することを意味する。ここで必要な幾何変換は透視変換であり、この変換に必要なパラメータは、実カメラと仮想カメラのパン角、チルト角、ロール角、3次元位置、画角および被写体の3次元位置である。

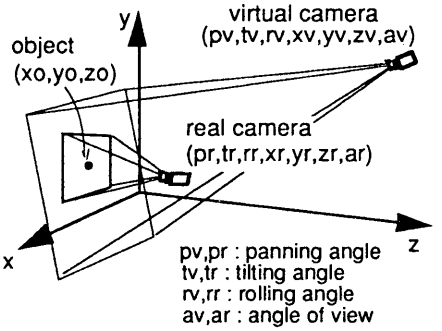


図2　被写体と実カメラ、仮想カメラの関係

Figure 2. Relationship among object, real camera and virtual camera

2．3　システムの構成

系統図を図3に示す。ユーザーが与えたカメラデータをもとにMCカメラおよび被写体回転台を駆動する。被写体回転台は等価的に被写体回り360度の撮影を行うためのものである。このとき、MCカメラおよび被写体回転台に取付けられた検出器により得られる実カメラデータと、ユーザーが与えた仮想カメラデータから計算された幾何変換量に従い、映像処理装置によりカメラ映像が処理される。なお、被写体は移動しないものとしている。また、図1（a）の撮影から（b）への移行をスムーズに行うため、MCカメラが可動限界に近づくにつれカメラの動きを徐々に遅くするような制御を行なう。前述した仮想撮影による映像処理により、図1（a）の撮影から（b）への移行自体は自動的に行なわれる。一方、背景セット映像は、仮想カメラデータをそのまま用いてリアルタイムCGを生成する。両者を合成することで完成映像が得られる。

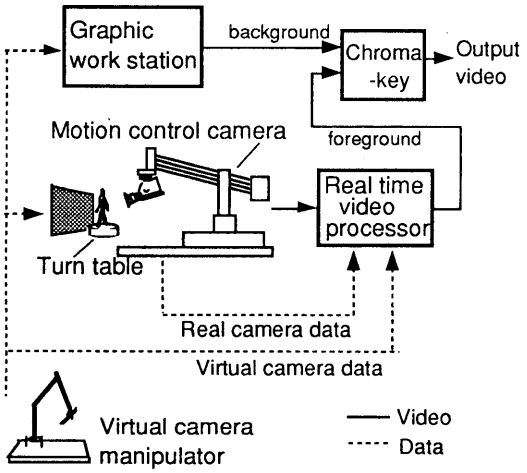


図3　システムの構成

Figure 3. Configuration of the system

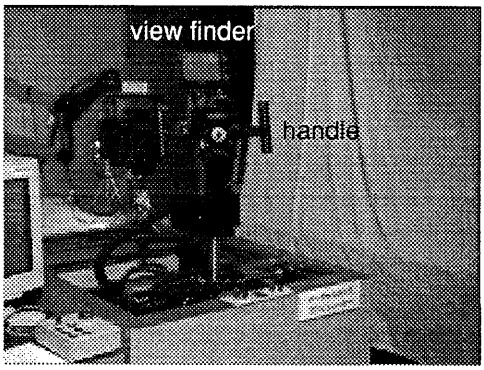


写真1　仮想カメラ操作部

Photo 1. Virtual camera manipulator

2．4　仮想カメラによる撮影

写真1に、仮想カメラ操作部を示す。この装置では、ハンドルの移動と回転によって仮想カメラの3次元位置とパン角、チルト角、ロール角を、手元のスイッチによってズーム、フォーカスのパラメータを調整す

る。また、これらの操作は通常のカメラと同じように、ビューファインダーのリアルタイム出力映像（写真2）を見ながら行うことができる。この仮想カメラによって、あたかも実世界の情景を「見る」ように、合成映像を制作することが可能になった。

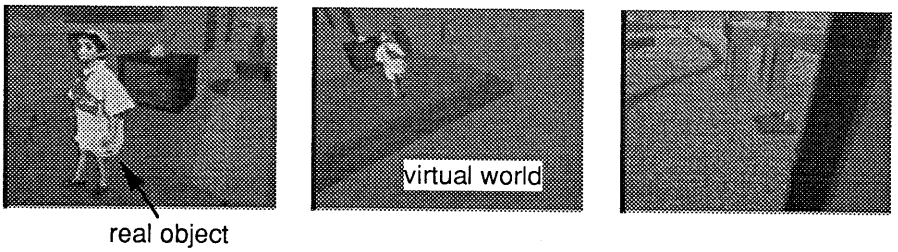


写真2　出力映像例

Photo 2. Output images of the system

3　見る人の視線を手がかりに　ー　インタラクティブ番組

いわゆるインタラクティブテレビとして語られるものは、テレビに双方向ネットワークとセットトップボックスを接続し、VOD(Video On Demand)やテレショッピングなどのサービスを受ける仕組みを指すことが多い。この場合のインタラクションとは、テレビのリモコンの形態のデバイスを駆使して、画面上のメニューから選択肢を選ぶといった比較的単純なものであり、映像の内容自体を変えていくような効果はない。もっと映像内容に対する関わりの大きいインタラクションは、いわゆるビデオゲームの中で実現されている。この場合には、ゲーム機に付属した比較的単純な操作デバイスを使って、ゲーム中のキャラクターやオブジェクトをリアルタイムに操る。この操作に従って、映像が変化していく。子どもにも簡単に使える操作デバイスのデザインと、CGなどのテクニックを駆使した立体的な画面作りにより、VRとでも呼べるような没入感をプレーヤーに催させることもある。

一方で、先に述べたように鑑賞するために作られた作品としての映像がある。この場合には、作家は見る人に映像で語りかけることになる。ただし、この語りかけは一方的で、見る人がどんな反応を示そうが、これとは関係なく進行していく。ここにインタラクションを持ち込むことができれば、この語りかけの幅を大きく広げることができる。例えば、見ている人がある話題に非常に興味を持っていることが分かれば、その話題についてどんどん内容を深めていったり、逆に退屈していることが分かれば、さっさと切り上げてしまうこともできる。このように、見る人の反応を知った上で、語り口を変えるような演出が可能になる。このためには、見る人の反応を感知する必要がある。しかも、見る人にとっては「見る」という自然な行為を妨げられないことが望ましい。以下では、この反応を感知する手段として、画面上の注視点の情報を使う映像提示システムと、試作したインタラクティブ番組について述べる。

3．1　インターフェースとしての視線追跡

コンピュータの分野では、次世代ユーザーインターフェースとしてNon-command-based interactionが期待されている[2]。現在のユーザーは、タスクを成し遂げるためのコンピュータのコマンドを選択し、これらをコマンドラインやメニューによってコンピュータに入力することによって仕事を進める。これに対し次世代のインタラクションでは、コンピュータがユーザーの状態を種々の方法で感知し、ユーザーのタスクが何なのかを把握して、そのタスクを成し遂げるために必要なプロセスを自動的に選択・実行する。このインタラ

クション環境では、ユーザーの状態の感知が重要なポイントになるが、この中でも視線追跡(Eye-tracking)は、ユーザーが何に興味を持っているかの情報を取得する手段として注目されている[3]。また、障害があって体を自由に動かせない人のためのインターフェースとして、ワードプロセッサなどへの応用も試みられている[4]。

今回は、映像を見る行為と連続的に繋がるインターフェースとして、見る人が何も着用しない型の視線追跡装置を使ったインタラクティブ映像提示システムを構成した。

3．2　システム構成

インタラクティブ映像提示システムの構成を図4に示す。

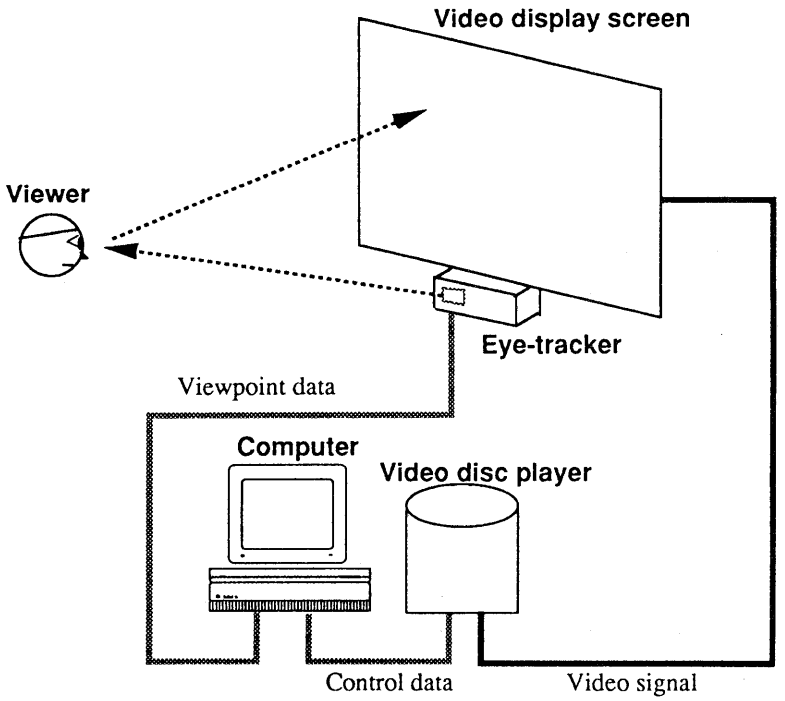


図4 システムの構成

Figure 4. Configuration of the system

視線追跡装置は、Viewerに近赤外の光線を照射し、ビデオカメラでとらえた眼の画像から、画像処理の手法で網膜と角膜表面の反射を分離し、この2つの情報から視線の情報を計算する。スクリーン上の複数の基準点によって視線のデータを較正しておくことにより、Viewerがスクリーンのどこを見ているかの情報（視線情報）が、リアルタイムで得られる。また、ミラーの自動制御により、15cmX15cm程度の範囲内でのViewerの頭の2次元的な動きには追随することができる。視線情報はコンピュータに入力され、コンピュータは視点情報に従って表示するビデオディスクの映像フレームを制御する。ビデオディスクは、映像をランダムに再生する機能を持っているため、Viewerには通常のスムーズな映像の流れが提示される。

3．3　インタラクティブ番組の構成

図5に、上記のシステムで提示する番組を構成するためのシナリオの構造を示す。

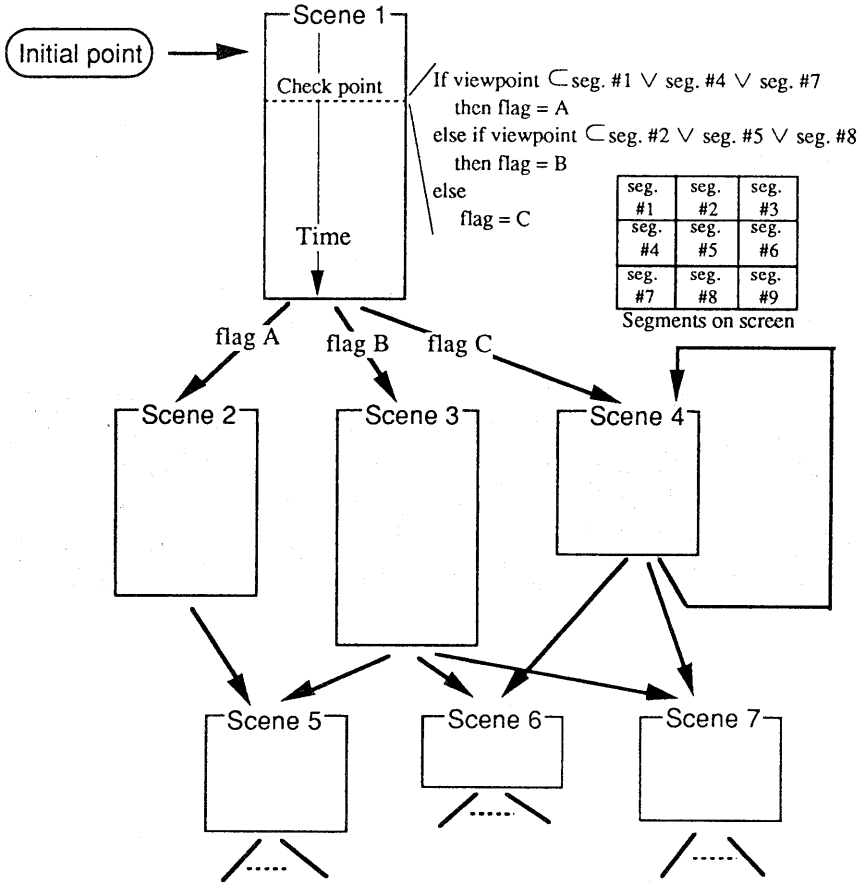


図5　インタラクティブ番組のシナリオの構造

Figure 5. Structure of a scenario of the interactive video

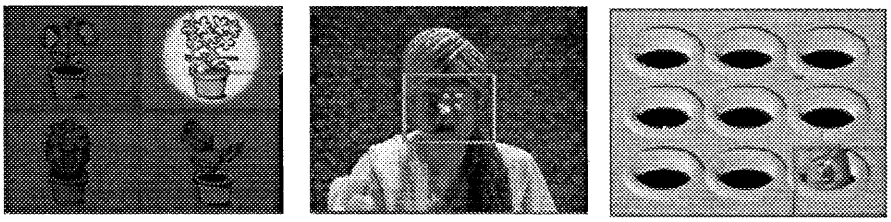
通常の番組のシナリオが一つの時間軸に沿った映像の展開を規定しているのに対し、このシナリオでは複数の時間軸や場合によってはループの存在も許す複雑な構造を持つ。

図6に、今回制作した番組の画面例を示す。今回は、画面全体を3 x 3の9領域または2 x 2の4領域に分割して、どの領域に視点が含まれるかの情報を番組の進行の制御に利用した。図6では画面中の矩形で、現在視点が含まれている領域を示している。Viewerの見る画面には、この矩形の表示は現れない。

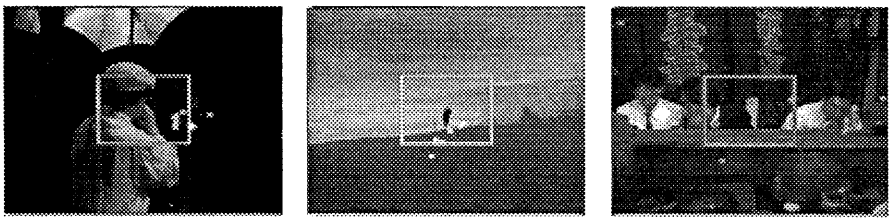
今回はゲームとドラマの2つのカテゴリーの番組を制作した。ゲームでは、キャスターの問いかけに応じて表示された選択肢を視線で選ぶクイズ形式のもの、眼をそらしたら負けの”にらめっこ”、視線を合わせることで射止める”もぐらたたき”などがある。このうちで、最初のクイズ形式のものは、現在のテレビ番組とよく似ているが、選んだ答えによって正解、不正解の表示がなされるだけでなく、以降の出題も変わってくるような仕掛けがなされている。他の2つは、いわゆるビデオゲームのタイプであるが、”にらめっこ”は実写映像が使われている。

以上のゲームは、Viewerが視線を感知されていることを意識している状態にあった。これに対し、ドラマでは選択肢はViewerには隠されており、いつ何を見たからこのような展開になったかはわからない。演出家が視点を誘導するような画面作りをしたり、それを逆手にとった仕掛けを作ったり、今までにないジャンル

の映像作品としての可能性は広いと考えられる。今回の作品はテーマが「悪夢」であり、ちょうど映像のWeb＝クモの巣の中をさまよい歩くような体験の提供を意図している。



Game



Drama

図６　制作したインタラクティブ番組

Figure 6. Images of the produced interactive video

4　まとめ

本稿では、映像メディアにおける「見る」インタフェースに関する２つの試みを紹介した。これらを、９５年６月のＮＨＫ放送技術研究所の一般公開で、展示する機会をえた。仮想カメラは、映像制作の新たなツールとして、実際に現場で映像制作に携わる人々の関心を呼んだ。視線追跡を使ったインタラクティブ番組は、見る人毎に較正をやり直さなければならないことや、頭を大きく動かすと追跡ができなくなるなどの技術的な問題はあるものの、「おもしろい」という評価は多く得られた。技術的な問題もさることながら、どのような演出が考えられるのかについて、まだまだ開拓の余地がある。このような、新しいインタラクションの形態の技術的開発とうまくリンクした内容（コンテンツ）の開発が、映像メディアの今後の発展にとっては重要であろう。

参考文献

[1] 林、福井、山内、長谷波：「番組人体２におけるハイビジョン映像合成技術」、第３回ハイビジョン研究会、2-2,(1993)

[2] Nielsen, J. : Non-command user interface. Communications of the ACM, 36(4), p.83-99 (1993)

[3] Jacob, R.J. : What You Look at is What You Get : Eye Movement-Based Interaction Techniques. CHI '90 Conference Proceedings. p.11-18 (1990)

[4] Fery, L.A., Preston White, J.R. and Hutchinson, T.E. : Eye-gaze word processing. IEEE Trans. on System, man and Cybernetics, 20(4), p.944-950 (1990)