2004－HI－111　(16)
2004／11／12

視線情報を利用した番組選択インタフェースの開発

小峯一晃[†]　澤畠康仁[†]　後藤淳[†]　小早川健[†]　浦谷則好[†]

頭部の自由な動きを許容する眼球運動測定装置を開発し，視線と音声対話でテレビの番組選択操作が可能なユーザ・インタフェースを試作した．画面上に複数の番組が同時に表示されるマルチ画面において，出力される番組の音声を注視点付近に表示されているものに切り替えるなど，番組内容確認の操作を支援する機能を組み込んだ．さらに，発話された「これ」などの指示代名詞に対し，その照応関係を注視点の位置から同定している．これらの機能により，音声対話インタフェース特有の煩わしさを部分的に解消するとともに，より自然な表現を用いた対話によって番組選択操作を行うことが可能になった．

A Development of User Interface for TV Program Selection Using Gaze Information

Kazuteru Komine[†], Yasuhito Sawahata[†], Jun Goto[†], Takeshi Kobayakawa[†]
and Noriyoshi Uratani[†]

We have made a prototype of easy-to-use speech dialogue interface for TV program selection using eye gaze information. A newly developed gaze tracking system permitting the users to move their head freely was applied in it. The viewers can switch the program sound of the multiple program thumbnails displayed on the screen, without any operation except eye movements. They can also utter demonstrative pronoun while gazing the thumbnail to select one of the programs.

This system enables the viewers to utter commands naturally and to select a displayed object easily by detecting their fixation point on a TV screen and resolving referent of demonstrative pronouns.

1 はじめに

地上波放送・衛星放送が共にデジタル化され，デジタル放送は薄型テレビ, DVD 録画機などのデジタル家電とともに，いよいよ普及の段階に入っていると言えよう．デジタル放送では多種多様なサービスが提供され，豊富な情報にアクセスできるようになった一方で，これらを享受するための受信機の操作は複雑になる傾向がある．高齢者を含む誰もが簡単に操作できるユーザ・インタフェースが望まれている．

我々は，テレビ画面上の GUI 操作が容易なリモコン[1]やデータ放送のユーザビリティ[2]などを実験により検証し，デジタル放送用の使いやすいリモコンの開発を進めてきた．[3]

†NHK 放送技術研究所
NHK Science and Technical Research Laboratories

その一方で，テレビがさらに高度化した際に想定される様々な情報へのアクセスを誰もが簡単に行えるように，音声対話を用いたユーザ・インタフェースの検討も進めている．[4]

音声対話インタフェースは，リモコン本体が要らない，検索時などに自由なキーワードが簡単に入力可能である，などの利点がある反面，認識率や曖昧な表現などの原因でユーザの意図が正確に把握できないという課題がある．そこで，我々はユーザの操作意図を正確に把握するために，音声認識で得られる言語情報以外の様々な非言語情報を利用し，マルチモーダル情報による頑健な意図推定手法の確立をめざしている．[5][6]

音声対話との併用が効果的な非言語情報として，視線情報が考えられる．視線は発話のタイミングや意志・感情などの意図を表出するため，意図推定に有効なモダリティである．さらに，テレビやPC のような情報を表示する画面を有する装置においては，注目している画面上のオブジェクトを取得するためユーザ・インタフェースとしても有用である．

視線情報を機器のユーザ・インタフェースに利用する研究事例として，PC 上のウィンドウ操作を視線によって行う例[7]や音声と組み合わせてエージェントとのコミュニケーションに利用する例[8]などが報告されている．また，視線からユーザの興味のある場所を推定し，表示を変えるなどの研究例もある[9]．

テレビは映像の視聴を主な目的とするメディアであることから，視線は自然な形で利用できるモダリティであると考えられるが，これまでのところテレビの操作に利用した例は少ない．

今回，テレビ視聴時にユーザが表出するマルチモーダル情報からユーザの意図を推定する手法を検討するためのなプラットフォームとして，視線情報と言語情報を利用したテレビ用ユーザ・インタフェースを開発した．また，テレビ視聴時のリラックスした状態での視線測定を想定し，頭部を固定せずに測定可能な視線測定装置を新たに開発した．

実際の操作性を検証するため，地上／BS デジタル放送の多チャンネル環境における番組選択操作を視線情報によって支援する機能を実装した．本報告では，視線と音声対話による番組選択操作のモデル，および開発したシステムの詳細について述べる．

2 視線情報による番組選択の操作モデル

従来のアナログ放送では，視聴できるチャンネル数が少なかったため，チャンネルを順次切り替えて内容を確認しながら視聴する番組を選択する「ザッピング」が主な操作モデルであった．

それに対し，デジタル放送，CATV などの多チャンネル環境では，ザッピングによる選択が非効率的になってしまうため，何らかの番組属性を手がかりに視聴する番組を選択する機会が多くなると考えられる．その際，番組の属性を表現するキーワードから候補番組を絞り込み，候補番組の内容を吟味しながら効率良く視聴番組を選択できる操作方法が望まれる．

我々が開発した音声対話インタフェース[4]では，ジャンル名や出演者名，番組名の一部などのキーワードを発話することにより，番組を検索する機能を実装しており，選択候補番組を効率的に抽出することが可能である．しかしながら，選択候補番組の中から視聴する番組を選ぶ際には，リスト表示された候補番組を一つ一つ選択して番組内容を確認しながら，最終的に一つの番組を決定する必要があった．

これらの煩わしさを解決する操作方法として，視線情報を用いた次のような操作モデルを提案する．

2.1 選択候補番組の表示

番組名等によるリスト表示（GUI）の代わりに，

複数番組の動画を同時に表示する複数画面表示（以下，6 画面表示）で候補番組を表示する．図 1 にその一例を示す．これにより，一覧性を向上させるとともに，次節以降に述べる方法により，その中から視聴する番組を 1 つ選ぶ操作を視線情報に基づいて支援する．なお，同時に表示する番組の数は直接記憶のマジックナンバー（7 ± 2）[10] を考慮して今回は 6 番組とした．以下に視線情報による操作支援の詳細を述べる．

2.2 番組音声の切替

6 画面表示された個々の番組の内容を確認する作業を視線で支援する．興味のある部分に視線を向ける視聴者の自然な動作を利用して，6 つの番組のうち注視している番組の音声を出力するとともに，テキストによる番組概要表示を行う機能を実装した．図 1 では右上の番組について番組内容がテキスト表示されている．視線の移動とともにフォーカス（緑色の枠）が移動し，番組の音声と内容表示がフォーカスの設定されている番組に切り替わる．なお，操作時点で放送されていない番組については番組の映像ではなく，番組のタイトル，放送チャンネル，放送時間を記した静止画を表示した（例：図 1 の右下）．

2.3 候補番組の交換

検索のキーワードによっては，絞り込まれた候補番組の数が同時に表示可能な 6 番組以上となることがあるため，表示されている番組以外の候補を表示する機能が必要となる．本操作モデルでは候補から外したい番組を注視しながら「この番組を交換して」「これ要らない」などの指示代名詞を含む発話によって，次の候補を表示することを可能にしている．この場合，指示代名詞の指示対象は視線情報で同定する．

また，「全部交換」などの発話で，全ての候補を入れ替える機能もある．

2.4 視聴番組の選択

上述の番組内容確認操作により，視聴する番組を決定する際には「これを見せて」などの指示代名詞を含む発話によって行う．これにより，フォーカスが設定されている番組が画面全体に表示され,通常の番組視聴状態になる．

図 1　候補番組が複数表示される画面の例

3 開発したシステムの構成

前述の番組選択操作モデルを実装するためのプラットフォームとして，視線と音声対話による操作が可能なテレビ用ユーザ・インタフェースシステムを試作した．本システムは主に音声対話処理部と視線情報処理部，および機器制御部の 3 つの部分に大別される．システム構成を図 2 に示す．視線情報以外の非言語情報を利用することも想定して，分散処理，自律動作，高拡張性を保てるような構成とした．

以下に各処理部の詳細を述べる．

3.1 音声対話処理部

音声認識，形態素解析，テンプレート比較処理を行い，音声合成や CG の応答を生成するモジュールである．また，放送波やインターネットから番組情報を検索する機能も含まれている（詳細は文献[4]を参照）．キーワードによって検索された候補番組のリストを番組情報とともに視線情報処理部へ送信する．これらの候補番組が 6 画面表示されている状況でユーザが発話した内容については，形態素解析した結果を視線情報処理部に送信する．形態素中に指示代名詞が含まれ，言語処理では対象同定ができない場合は，指示対象同定要求を視線情報処理部へ送信する．視線情報処理部で同定

可能な場合は，指示対象が返送されてくる．その結果に基づいてテンプレート比較処理を行う．

3.2 視線情報処理部

①視線測定モジュール

試作したシステムでは，非接触型の視線測定装置を用いてユーザの視線を捉えている．既存の非接触型測定装置が具えている眼球追従機構では，頭部の速い動きには追従しきれない．そのため，今回新たにユーザの頭部周辺を撮影するカメラを併用し，頭部の速い動きにも追従できるように既存の装置を改修した．詳細については 4 章で述べる．

②オブジェクト検出モジュール

視線測定装置から出力される注視座標と表示画面の情報を利用して，注視している画面上のオブジェクト（以下，注視 OBJ）を検出する．注視座標から注視オブジェクトへの変換については0 で詳述する．検出された注視 OBJ が現在選択されている（フォーカスが設定されている）オブジェクトと異なる場合は注視 OBJ 変更を示すイベントを意図推定モジュールに送信する．同時に注視 OBJ をイベント発生時刻とともに履歴として蓄積する．

一方，音声対話処理部から発話情報（形態素解析の結果）が送信されてきた場合，意図推定モジュールから発話開始時刻における注視 OBJ を問い合わせるコマンドが送信されてくる．その時刻における注視 OBJ を履歴から抽出し，意図推定モジュールに返送する．

③意図推定モジュール

音声対話処理部とオブジェクト検出モジュールから送られてくる情報を統合し，指示語の同定や視線によるフォーカス移動の処理を行うモジュールである．

音声対話処理部から番組情報リストを受信した場合はリストにしたがって 6 画面表示を行うよう機器制御部にメッセージを送信する．また，注視オブジェクト変更のイベントを受信した際には，フォーカス移動のメッセージを機器制御部へ送信し，番組内容表示や出力音声の切替を行う．

対話処理部から送信されてくる発話の形態素解析結果に指示代名詞が含まれていた場合，発話からは操作意図が定まらない．②のモジュールに発話開始時刻をパラメータとする注視 OBJ 要求を

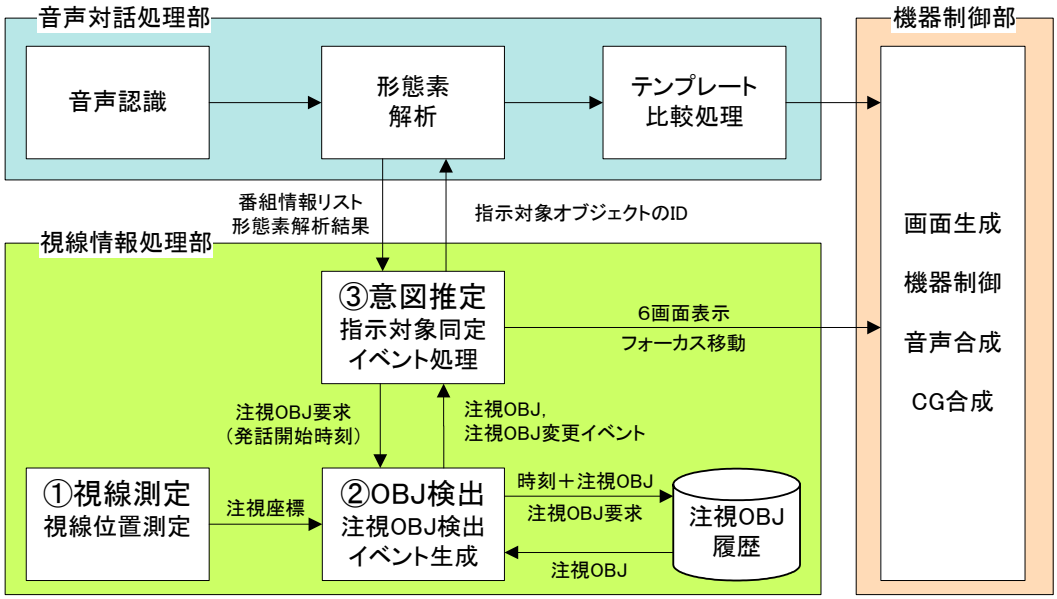


図 2　開発したシステムの構成図

送信することにより，発話開始時点における注視 OBJ が返送され，それによって指示対象の同定を行う．その際，発話開始時刻は，発話された文字数から推定している．同定した指示対象のオブジェクトを番組情報に記述されている ID に変換して対話処理部へ返送する．

3.3 機器制御部

意図推定モジュールから送信される操作意図や対話処理部で生成される応答に基づいて，チャンネルの切替や画面の表示制御を行う．ユーザへの応答は画面上のフォーカス移動，合成音声，CG によって行っている．

4 頭部の動きを許容する視線測定装置の試作

4.1 テレビ視聴環境における視線測定の所要条件

テレビ視聴時にユーザの視線を測定するためには，テレビの性質上，リラックスした状態で測定できることが望ましい．したがって，非拘束性が重要な所要条件となる．この場合，頭を動かさずに（眼球位置を固定したままで）視聴することは考えられないため，眼球位置を追跡するための機構が必要となる．

4.2 測定装置試作の方針

本システムでは，2 台のカメラを用いて視線測定を行う．第 1 のカメラは，眼球のアップを撮影し，第 2 のカメラは，ユーザの頭部周辺領域を撮影する．第 2 のカメラの情報に基づき，第 1 のカメラを制御することで，ユーザの頭部の動きを許容する視線測定装置を実現した．（図 3）

また，測定された視線データ（座標値）は，生理的な微動によるデータのばらつきや，測定時のノイズが含まれており，視線測定装置の出力をそのままポインタとして利用するのは適切ではない．本システムでは，注目オブジェクトの検出アルゴリズムを実装することで，操作時の違和感を減らす工夫をしている．

以下，これらの詳細について述べる．

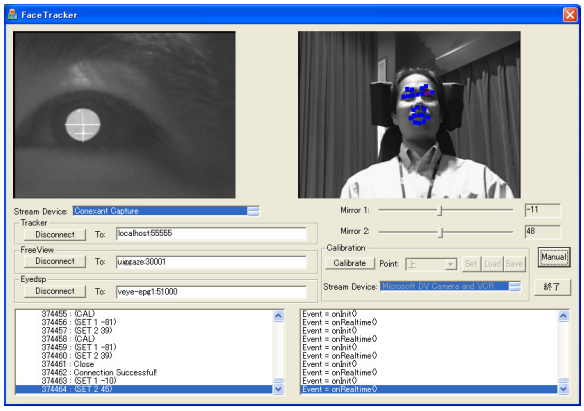


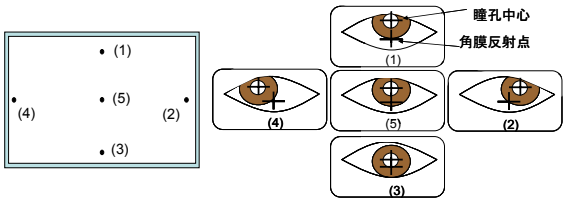
図 3　視線測定と眼球追跡

4.3 視線測定の原理

非拘束性を実現した視線測定手法として，眼球に近赤外光を当て，角膜表面からの反射光（プルキニエ像）と瞳孔中心位置を画像処理により求め，それらの相対位置の変化により視線方向を得る手法が知られている．視線位置が変わると，瞳孔中心とプルキニエ像の位置は図 4 のように変化する．

詳しい方法はここでは省略するが，これらの位置関係の変化を利用することで，視線位置を測定することができる．

しかしながら，この測定原理では，キャリブレーション位置と実際の測定位置が異なる場合は，系統的な誤差が生ずる．この問題に対しては，大野らの方法[11]を利用することで対応が可能である．



(a)画面上の点　　(b)画面上の点を見た際の眼球の様子

図 4　瞳孔中心とプルキニエ像の関係

4.4 眼球位置の追跡

最初に眼球の位置を見つける際や，第 1 のカメラの方向制御（既存の測定装置では，カメラレンズの前に設置したミラーの角度によって撮影方向

を制御している）が追いつかない程の早い動きがある場合は，ユーザの周辺領域を捉える第 2 のカメラにより目の位置を検出し，眼球の追跡を開始/再開することができる．このためには,顔の特徴点から適切なミラー角度を計算する際に，事前のキャリブレーションが必要である．

図 5(b)のような，3 点が描かれた板を第 2 のカメラで撮影し，それぞれの点を第 1 のカメラが捉えられるようにミラーを調節する．3 点の座標とミラー角度との組の関係が，線形であると仮定することで，第 1 のカメラが第 2 のカメラ内の任意の点を捉えるミラー角度を計算することができる．キャリブレーションの手順は次のとおりである．

図 5(a)のように，点 1 が画像の中心に来るようにミラーを動かす．このときのミラーの角度を，(m_{1h}, m_{1v}) とし，図 5(b)内の点 1 の座標値を (x_1, y_1) とする．同様の作業から，点 2，点 3 からも，(m_{2h}, m_{2v})，(m_{3h}, m_{3v})，(x_2, y_2)，(x_3, y_3) が得られる．これらの点を x 方向，y 方向について線形補間することにより，第 2 のカメラで捉えた画像内の任意の点 (x, y) を第 1 のカメラで捉えるためのミラーの角度 (m_h, m_v) は，次式で与えられる．

$$m_h = \frac{m_{3h}x_2 - m_{2h}x_3}{x_2 - x_3} + \frac{m_{2h} - m_{3h}}{x_2 - x_3}x \quad (1)$$

$$m_v = \frac{m_{1v}y_2 - m_{2v}y_1}{y_2 - y_1} + \frac{m_{2v} - m_{1v}}{y_2 - y_1}y \quad (2)$$

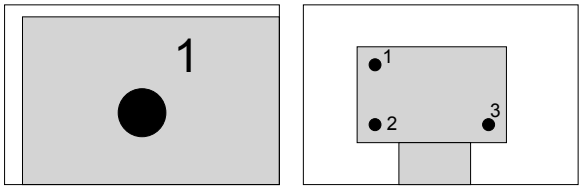
但し，ユーザがカメラからの距離方向に移動するような状況では，式(1)，(2)は不適切になる．距離方向への移動に対応するためには，フォーカスなどによる距離測定機構と距離による幾何学的な補正が必要となる．

ところで，ミラーが動いている時は，撮影している眼球画像が乱れるため，視線の測定ができない．ユーザの動きに対応しつつ，ミラーを動かす頻度を最小限に抑える必要がある．第 1 のカメラ単体による追跡は，眼球画像を捉えているか否かに基づきミラーを動かすため，ミラーを動かす頻度は最小であるが，ユーザの早い動きには対応できない．一方で，第 2 のカメラを併用した場合の追跡は，ユーザの早い動きには対応可能だが，眼球画像を捉えているかどうかを考慮していない．

これらを鑑み，できるかぎり第 1 のカメラ単体による追跡方法を利用し，その追跡方法では対応しきれない動きがあるときのみ，第 2 のカメラを併用する方法を採用する．具体的には，次のようなタイミングで切り替える．

まず, 単位時間あたりの視線データの出力数（取得レート）をチェックする．視線の測定が適切に行われている場合は，取得レートは約 30[sample/sec]であるが，そうでないときは，取得レートが著しく低下するという特徴がある．

本システムでは，しきい値を 10[sample/sec]とし，取得レートがこの値を下回ったとき，眼球追跡の方法を，第 2 のカメラを併用する追跡方法に切り替えている．これにより，ユーザの早い動きに対応しつつ，最小頻度のミラー操作で視線測定が可能になる．



(a)第 1 カメラの画像　　(b)第 2 カメラの画像

図 5　キャリブレーション

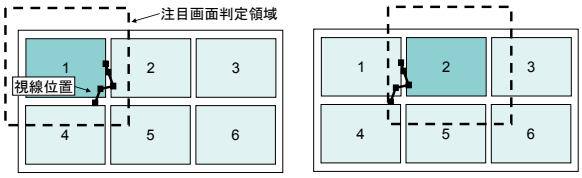
4.5 注目画面の設定方法

視線測定装置により得た測定値（座標値）は，測定時におこるノイズや，人間の生理現象に基づく不随意な運動により，ユーザが一点を凝視しているつもりでも，ばらつきをもって測定される．そのため，測定値を直接ポインタとして利用すると，システムの使い勝手を低下させてしまう．そこで，測定データのスムージングと，注目画面を決定する領域のサイズを動的に変更する処理によ

り，使い勝手の向上を図った．

測定された(x, y)の座標値をバッファし，ソートを行い，バッファ内の中央値(メジアン)を現在の視線の位置とすることで，スムージングをしている．バッファの長さは，経験的に 10 としている．新しい測定値が到着したときは，バッファ内でもっとも古い測定値が破棄され，新測定値はバッファ内の適切な位置に挿入される．これにより，ノイズにより視線位置が大きく外れる現象が防げる．

それでもなお，不随意な運動による測定値のばらつきは残る．これに対応するため，注目画面判定領域を図 6 のようにとった．図中の破線で示された注視画面判定領域を脱したときに，注視画面が変わったと判定する．これにより，不随意な運動による視線の細かい変動により，注視画面が頻繁に切り替わるという現象を除くことができる．



(a)子画面 1 を注目時　　(b)子画面 2 を注目時

図 6　注目画面判定領域

5 実装

実際の操作性を確認するためにデジタル放送の番組を選択できる機能を実装した．地上／BS デジタルチューナ（松下電器（株）TU-MHD500）6 台とスキャンコンバータ（アストロデザイン（株）MC-2004）2 台を用いて 6 画面表示（図 1）を生成した．画面の制御は，システムで推定された操作意図に基づいて，チューナのチャンネルとスキャンコンバータの制御を行うことにより実現している．システムの各モジュールの開発にあたっては，マルチモーダルによる意図推定のプラットフォームとしての利用を考慮し，各モジュール間の連携はテキストベースのソケット通信によって行っている．これにより，モダリティの増減に柔軟に対応するスケーラビリティを確保するとともに，各モジュールの自律した動作が可能となる．

また，視線測定装置として，竹井機器工業（株）の FreeView DTS を利用した．30[sample/sec]での視線データの出力ができる．この装置には，赤外線の光源とカメラ（第 1 のカメラ）およびミラーが搭載されている．第 2 のカメラとしては，Canon VC-C4（雲台一体型カメラ）を利用した．顔の特徴点を抽出するソフトウェアは，N-Vision（株）の SDK を利用した．顔の特徴点は，30[fps]で追跡が可能である．

なお，ユーザ・インタフェースシステム，視線測定装置ともにソフトウェアの実装には Visual C++を用いた．

6 動作検証

50 インチの PDP に映像および操作画面（6 画面表示）の表示を行い，視距離約 2m（約 3H，H は画面の高さ）の位置で視聴および操作した．視線測定装置はユーザの眼球から約 1.2m の位置に設置した．

実際に BS／地上デジタル放送の番組選択を本システムで行ったところ，視線測定が可能なユーザでは所望の動作が可能であることを確認した．

また，視線測定装置については，頭部の動作によって眼球追跡が外れたときから，再度眼球の追跡を復帰するまでにかかる時間は，約 1[sec]であった．注視している子画面を決定するアルゴリズムについては，ユーザの内省報告ではおおむね好評を得ている．

これらの評価ついては今後，客観的な評価を得るための実験を行う予定である．

7 まとめ

視線情報を利用して音声対話によるテレビの操作を支援するシステムを試作した．視線情報をポインティング操作や指示代名詞の指示対象同定に

利用することにより，音声対話のみでは煩わしかったザッピング操作や画面上オブジェクトの選択操作を自然な発話で行うことが可能になった．

現在，試作したシステムの操作性を評価するため，評価実験を行い，その結果を解析中である．

今後は，本システムを利用して，視線の動きからユーザの興味や操作支援要求を推定する手法の検討についても行っていきたい．

参考文献

[1]小峯ほか：“テレビ画面上の GUI 操作環境における高齢者のリモコン操作性評価”，映像情報メディア学会論文誌，Vol.55，No.10，pp.1345-1352（2001）

[2]森田ほか：“高齢者におけるデータ放送コンテンツのユーザーインターフェース評価",ヒューマンインタフェース学会研究報告集，Vol.4，No.5，pp.75-80（2002）

[3]吉田ほか：“デジタル受信機のための少ボタン型リモコンによるヒューマンインタフェースの試作"，映像情報メディア学会年次大会予稿集,9-1（2004）

[4]J.Goto et.al. :”A Spoken Dialogue interface for TV Operations Based on Data Collected by Using WOZ Method”, IEICE Transactions on Information and Systems, Vol.E87-D, No.6, pp.1397-1404 (2004)

[5] 森田ほか：“視聴者の意図に基づいたテレビインタフェースの提案”, FIT2003，K-056，pp. 547-548 (2003)

[6]小峯ほか：“テレビ視聴者の操作意図を推定するためのマルチモーダルデータベースの枠組み”，FIT2003，K-057，pp. 549-550 (2003)

[7]大野：“視線を利用したウインドウ操作環境”，信学技報，HIP99-29，pp.17-24（1999）

[8]知野ほか：“Gaze To Talk ：メタコミュニケーション能力を持つ非言語メッセージ利用インタフェース”，インタラクション’98 論文集，pp.169-176（1998）

[9]Starker, I. et.al. : “A Gaze-Responsive Self-Disclosing Display, Poceedings of Conference on Human Factors in Computing System (CHI’90), pp.3-9 (1990)

[10]Miller, G. A., “The Magical Number Seven, Plus or Minus Two: Some limits on our capacity for processing information”, Psychological Review, 63（1956）

[11]大野ほか：“2 点補正による簡易キャリブレーションを実現した視線測定システム”，情報処理学会論文誌，Vol.44，No.4，pp.1136-1149（2003）