

K-074

## 視線情報を利用したテレビ用ユーザインタフェースの開発 Development of User Interface for TV Operation Using Gaze Information

小峯一晃\* 澤島康仁\* 後藤淳\* 小早川健\* 浦谷則好\*  
Kazuteru Komine Yasuhito Sawahata Jun Goto Takeshi Kobayakawa Noriyoshi Uratani

### 1. はじめに

複雑化するテレビの操作を誰もが簡単に使えるテレビ用ユーザインタフェースが望まれている。我々はその一つの解決策として音声対話インタフェースによるテレビの操作を提案している<sup>[1]</sup>。

音声対話による操作は多チャンネル視聴環境において番組を検索する際に、番組名や出演者名などの自由なキーワードを簡単に入力でき、インタフェースとして有効なモダリティである。その反面、画面上に表示されているメニューなどのオブジェクトを指示することが煩わしいという側面がある。また、操作を自然な音声対話で行う際には「この」「その」などの指示代名詞を用いた表現が頻繁に発声されると考えられる。このような音声対話インタフェースの課題を解決する手段として、ユーザが表出するマルチモーダル情報を利用することを検討している<sup>[2]</sup>。なかでも視線は意図を表現する有効なモダリティであり、特に情報を表示する画面を有する装置においてはポインティングデバイスとしても有用であると考えられる。

視線情報を機器のヒューマン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<sup>[3]</sup>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<sup>[4]</sup>などが報告されている。

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推測されるが、これまでのところテレビの操作に利用した例は報告されていない。

今回、視線情報を利用して画面上のポインティング操作や発話された指示代名詞の指示対象の同定を行うテレビ用ユーザインタフェースを試作した。本報告では試作システムの概要とテレビ操作における視線の利用例について述べる。

### 2. 視線を利用した操作

多チャンネル環境では、チャンネルを順次替えて番組を選ぶ従来のザッピング操作に比べ、何らかの番組属性を手がかりに多くの番組の中から1つの番組を選択する操作を行う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、候補番組の内容を確認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。

従来の音声対話システム<sup>[1]</sup>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により入力し、検索結果の番組選択候補を GUI によるリストで表示していた。候補番組の内容を視聴して確認するためには番組の検索・選択・内容確認を繰り返す必要があった。

本システムではリスト表示の代わりに、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）で候補番組を表示して番組内容の確認を容易にするとともに、

その中から目的の番組を1つ選ぶ際の操作に対して視線で支援する機能を実装した。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（ $7 \pm 2$ ）<sup>[5]</sup>を考慮して今回は6番組とした。以下に詳細を述べる。

#### 2.1. 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。見たい部分に視線を向ける視聴者の自然な動作を利用して、6つの番組のうち視線の交わっ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に番組内容がテキスト表示された状態の画面を示す。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。テキストで表示する番組の概要はデジタル放送の EPG で利用されている情報（放送波で配信される番組配列情報）から抽出している。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）。

#### 2.2. 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。本システムでは候補から外す番組を注視しながら「この番組を交換して」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が可能である。この場合、指示代名詞の指示対象は視線情報で同定している。

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実装した。

#### 2.3. 視聴番組の選択

上述の番組内容確認操作により吟味した番組を視聴する際には「これを見せて」などの指示代名詞を含む発話によって、フォーカスが設定されている番組を画面全体に表示することができる。



図1 候補番組が複数表示される画面の例

\*NHK放送技術研究所

### 3. システム構成

前述の番組選択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。以下に各処理部の詳細を述べる。

#### 3.4. 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[1]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態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る場合は、指示対象が同定され、その結果が返送されてくる。それに基づいてテンプレート比較処理を行い、適切な応答を生成する。

#### 3.5. 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ユーザの視線位置を捉える方法として非接触型の視線測定装置(竹井機器製 FreeView)を用いている。本装置は赤外線カメラで撮影した眼球部の映像から画像処理によって視線を測定する方式の装置で、光軸から照射したLEDの角膜反射像(ブルキニエ像)の位置と瞳孔の中心位置とから視線方向を算出している。

装置内部で眼球部の動きを追従する制御を行っているが、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上半身を撮影したカメラを併用し、頭部の速い動きにも追従できるように改修した<sup>[6]</sup>。また、検出された視線の位置をポインティング位置としてそのまま利用すると眼球の不随意的動きを反映することとなり、意図していない動作になるため、出力を平滑化するためのフィルタリング処理を行っている(詳細は文献[6]を参照)。

##### 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視線位置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。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。

#### 3.6. 機器制御部

意図推定モジュール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。操作画面(図1)の生成には地上波デジタルチューナ6台と縮小した動画を表示するためのスキャンコンバータを用いている。

### 4. 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。

今後は試作したシステムの操作性評価を行い、視線利用の有効性を検証する。また、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を検討していきたい。

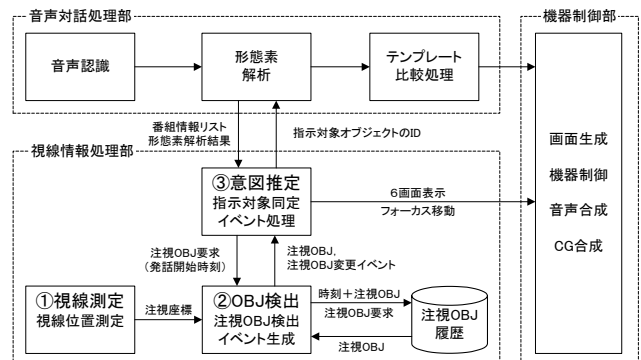


図2 開発システムの構成図

### 参考文献

- [1]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)
- [2]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)
- [3]大野: "視線を利用したウィンドウ操作環境", 信学技報, HIP99-29, pp.17-24 (1999)
- [4]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", インタラクシオン'98 論文集, pp.169-176 (1998)
- [5]Miller, G. A., "The Magical Number Seven, Plus or Minus Two: Some limits on our capacity for processing information", Psychological Review, 63 (1956)
- [6]澤島ほか: "頭部の自由な動きを許容する視線測定システム", FIT2004 (2004)