

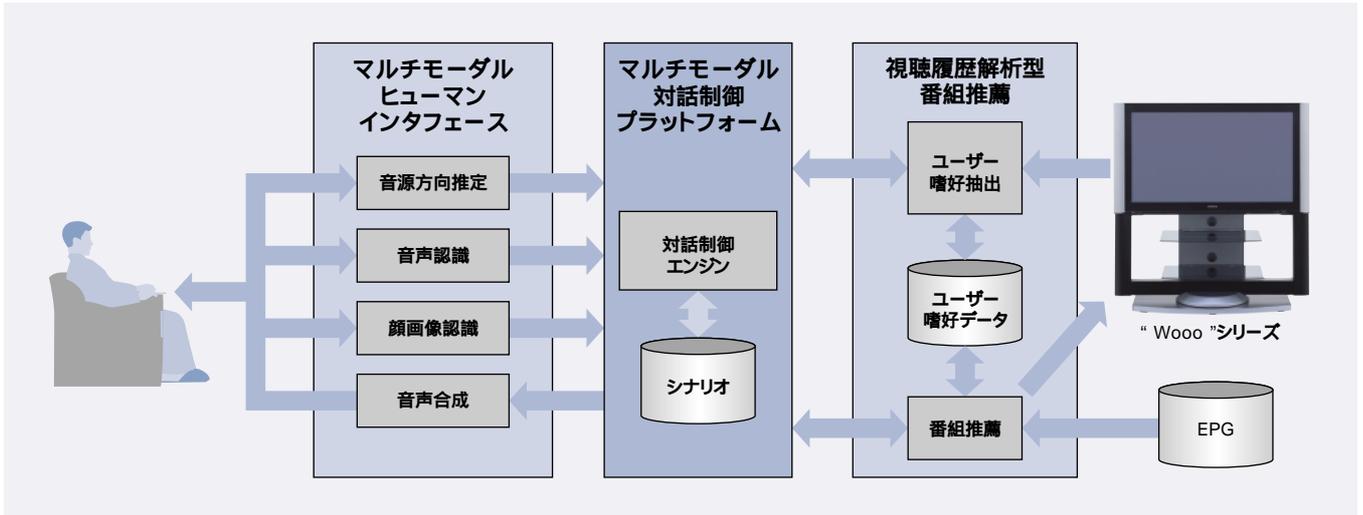
# マルチモーダル対話技術による 知的ユーザーインターフェース

## Intelligent User Interface Based on Multimodal Dialog Control for Audio-Visual Systems

新庄 広 Hiroshi Shinjō  
山口 宇唯 Ui Yamaguchi

天野 明雄 Akio Amano  
内部 こなぎ Konagi Uchibe

石橋 厚 Atsushi Ishibashi  
桑本 英樹 Hideki Kuwamoto



注:略語説明 EPG( Electronic Program Guide ), AV( Audio-Visual )

### マルチモーダル対話技術を用いた知的ユーザーインターフェースの構成

音声や画像など、さまざまな入出力方式を統合したマルチモーダル対話に基づいて、ユーザーの状況に応じたサービスを提供することができる知的ユーザーインターフェースを開発した。

AV機器の高機能化や放送の多チャンネル化に伴い、番組を「選びきれない・見きれない」という時代が近づきつつある。この問題に対応するため、日立製作所は、音声対話によってAV機器の操作を支援する「マルチモーダル対話技術による知的ユーザーインターフェース」を開発した。

マルチモーダル対話技術は、音声情報処理技術と画像情報処理技術、および視聴履歴解析に基づく番組推薦機能を統合し、ユーザーと対話を行う技術である。番組

推薦技術とは、キーワード入力などを必要とせずに、ユーザーの視聴履歴から、自動的に番組の好みを分析し、番組推薦や番組自動録画、番組選択などを行うものである。この知的ユーザーインターフェースは音声情報処理により、自然言語に近い会話でAV機器を操作でき、顔画像認識でユーザーを識別し、番組推薦などのサービスを、ユーザーごとに自動的に切り替えることができる。

## 1 はじめに

地上デジタル放送の開始に伴い、AV( Audio-Visual )パソコンやHDD/DVD( Hard Disc Drive/Digital Versatile Disc )レコーダが普及してきている。また、ハードディスクの容量は今後もますます増え続け、「何でも録(と)っておく」時代も近いと考えられる。さらに、家庭内で各種AV機器をネットワークで接続することによってホームサーバにするなど、AV機器も高機能化、複雑化している。このような状況の下で、AV機器、特にテレビの利用形態は、「好きな番組を「自主的に見る・録る」という状況から、「選びきれない・見きれない」という状況へ変

化しようとしている。

このような課題に対し、日立製作所は、音声合成・認識、画像認識、テキスト解析技術を統合したマルチモーダル対話制御により、AV機器の操作を支援する新しい知的ユーザーインターフェースを試作した。

ここでは、この知的ユーザーインターフェースのコンセプトとその要素技術について述べる。

## 2 知的ユーザーインターフェースのコンセプト

この知的ユーザーインターフェースは、自然言語を用いた対話形式による直感的な操作を通して、多数の放送

番組や録画番組の中から見たい番組を効率よく選択するなど、AV機器の操作を支援するものである。

知的ユーザーインターフェースは、音声認識技術、音声合成技術、顔画像認識技術、視聴履歴解析による番組推薦技術、およびマルチモーダル対話技術の要素技術を統合したシステムとして構成されている。

これらの要素技術により、以下のような機能を実現している。

(1) 親しみやすいインタラクション機能

音声認識、音声合成、画像認識など、複数の入出力手段を統合して制御するマルチモーダル対話技術により、人間との対話と同じようなユーザーインターフェースを実現する。

(2) 番組推薦機能

キーワード入力などを必要としないで、ユーザーの番組視聴履歴から好みの番組を自動的に学習し、番組推薦や自動録画を行う。

(3) 個人適応型サービス提供機能

AV機器の前に居るユーザーの顔からユーザーが誰であるかを自動的に判別し、ユーザーごとに最適なサービスを提供する。例えば、ユーザーに個人名で呼びかけたり、ユーザーの番組の好みに応じて録画番組や放送中の番組の視聴を音声対話で推薦する。また、いつも見ている番組の時間になると視聴を通知することもできる。

(4) 離席時の自動録画と後追い再生機能

ユーザーが放送番組の視聴中に所用で席を立った場合に、番組表示を一旦中断してその番組録画を開始する。その後、ユーザーが席に戻ってくると、本人識別を行い、中断したところから再生を始める。

以上の機能以外にも、さまざまなサービスの提供へ向けて拡張が可能である。

### 3 要素技術

#### 3.1 音声認識技術

複数のマイクを用いてマイクアレイを構成することにより、音や音声の到来方向を検知する機能や、妨害音の存在下で目的の音声だけに焦点を当てるといった音源分離機能を実現した。この方式では、8個のマイクを縦横の二次元状に配置することにより、方位角(水平方向)と仰角(垂直方向)を同時に検知し、音源分離ができる。

音源分離には周波数振り分けによって求めた目的音方向と妨害音方向の情報に基づいて適応フィルタの指向性および死角形成を制御する手法をとっており<sup>1)</sup>、最大20 dBの妨害音抑圧性能を達成している(図1参照)。

音源分離によって得られた音声波形はカーナビゲ-

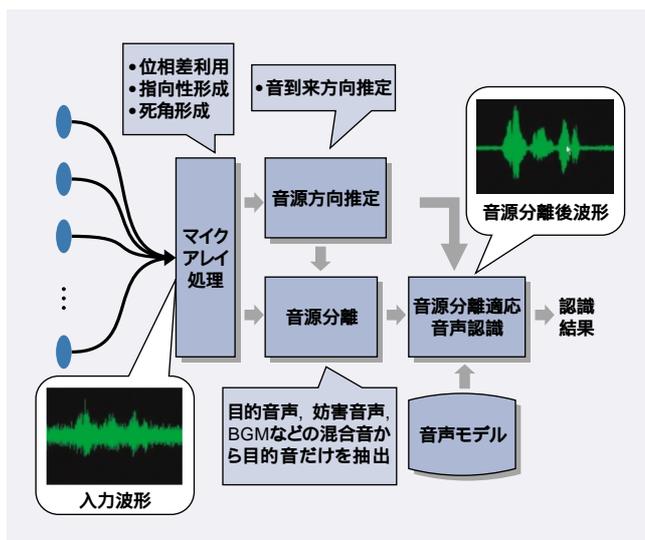


図1 音声認識処理の流れ  
マイクアレイに到来する音の位相差と振幅差を利用して音源方向を推定し、さらに、目的音に焦点を当てて抽出した音声に音声認識処理を施す。

ション向けに開発してきた音声認識エンジン<sup>2)</sup>に入力され、テレビチャンネルの選択をはじめとするAV機器制御用の音声コマンドを認識する(図1参照)。

#### 3.2 音声合成技術

テキストデータから肉声感の高い音声を合成できる、高品位知的音声合成技術を開発した。AV機器のユーザーインターフェースでは、ユーザーの名前や番組名など、あらかじめ登録できないことばを正確に発話するために、録音済みの音声を再生するのではなく、音声を合成する技術が必要である。

日立製作所は、独自の「調音コスト最小化方式」を開発し<sup>3)</sup>、最適な音声素片を効率よく選択し、選ばれた素片を滑らかに接続することにより肉声感が高く滑らかで、

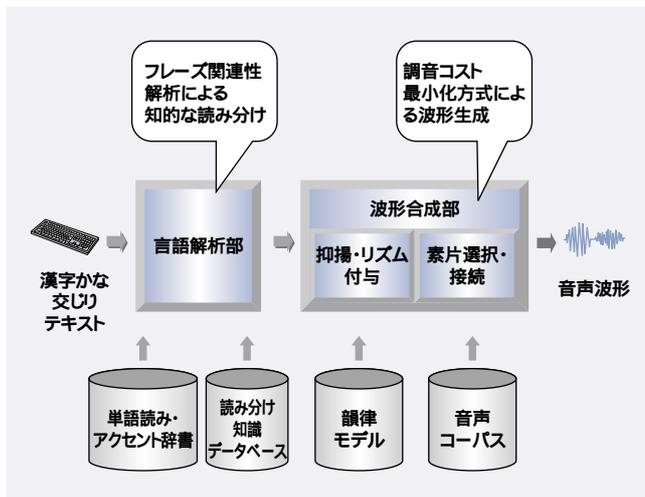


図2 テキスト音声合成処理の流れ  
漢字かな交じりテキストに対して、読みおよびアクセントを決定する「言語処理」と、音声コーパスから最適な素片を選択して、波形を生成する「波形合成処理」から成る。

多様な音声の合成を実現した(図2参照)。また,番組情報を音声で案内するために,EPG(電子番組表)データを正しく読み上げる「フレーズ関連性解析による読み判定技術」を開発した。複数の読み方がある単語であっても,単語間の関連性データと前後の文脈から読みを判定し,複雑な漢字かな交じり文の知的な読み分けを実現した。

### 3.3 顔画像認識技術

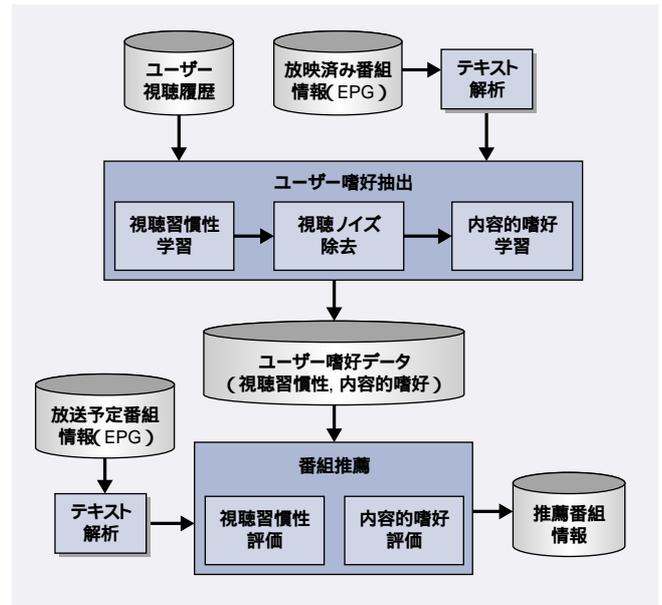
映像からユーザーの顔を検出し,個人を識別する技術を開発した。処理の概要は以下のとおりである。まず,顔検出用フィルタ群を用いて画像の中から顔領域を検出する。次に,顔領域の中から目・鼻・口などの顔器官(特徴点)を検出し,特徴点周辺の濃淡情報から顔の個人差を表現する特徴量を算出する。この特徴量を,あらかじめデータベースに登録済みの特徴量と比較することにより,ユーザーが誰であるかを識別する(図3参照)。

顔画像認識の技術課題としては,設置場所の移動や使用時間,逆光・斜光などの照明条件の変動,顔の角度の変化に伴う顔の見え方の違いを吸収することなどがあげられる。この課題に対し,正準判別分析などの統計的モデルを用いて精度向上を図っている。

この方式を,80人(8,000枚)の顔画像データベースで評価した。顔に対する照明方向は,水平方向が $-45 \sim 45$ 度,垂直方向が $0 \sim 60$ 度,顔角度は水平垂直とも $-10 \sim 10$ 度である。このデータベースでの識別率は,正面顔,水平方向に $\pm 10$ 度傾いた顔,垂直方向に $\pm 10$ 度傾いた顔のそれぞれに対して,98.0%,95.1%,96.9%であった。

### 3.4 視聴履歴解析による番組推薦技術

番組内容に対するユーザー嗜好(内容的嗜好)と,毎日・毎週,習慣的に番組を視聴する傾向(視聴習



注:略語説明 EPG( Electronic Program Guide; 電子番組表 )

図4 嗜好抽出・番組推薦機能の概要

内容的嗜好と視聴習慣性を学習し,学習結果を反映した番組を推薦をする。

慣性)を学習し,学習結果を反映した番組推薦を行う(図4参照)。

ユーザー嗜好抽出部は,EPGから取得できる番組タイトル,ジャンル,出演者,概要説明などについてテキストを解析して抽出したキーワード群と,ユーザーが視聴した番組の視聴時刻・頻度などの履歴を入力する。嗜好抽出の際,番組内容にかかわらず習慣だけで視聴しているニュースなどの番組(視聴ノイズ)による精度低下が問題となる。そのため,視聴習慣性学習結果に基づいて視聴ノイズを除去することにより,高精度な内容的嗜好学習を実現した。

番組推薦部では,放送予定番組について視聴習慣性の評価と内容的嗜好評価を行い,推薦の可否を判定する。

### 3.5 マルチモーダル対話制御プラットフォーム

ユーザーと各種AV機器やデータベースとの間を仲介するマルチモーダル対話制御プラットフォームを開発した。

このマルチモーダル対話制御プラットフォームは,ユーザーの要求を実行するための「シナリオ」をXML( Extensible Markup Language )を用いて記述している。このシナリオを適宜,追加・変更することにより,多種多様な機能を実行できる。この研究では,上述の音声認識・合成技術,顔画像認識技術,番組推薦技術を統合し,AV機器の操作に特化したシナリオを作成することにより,知的なインタフェースを実現した。



図3 顔画像認識結果の例

逆光などの悪条件下でも,顔領域や顔器官(特徴点)を正しく検出し,人物を識別する。

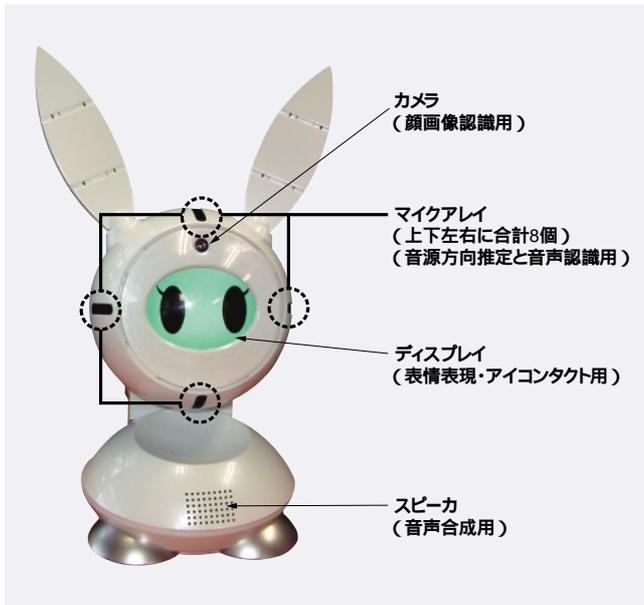


図5 知的ユーザーインターフェースの試作マスコットの概略構成  
音声認識用の8個のマイクと、顔画像認識用のカメラ、音声合成用スピーカ、表情をあらわすディスプレイなどを装備する。

## 4 試作

上述した技術を統合した知的ユーザーインターフェースを、うさぎ形の外観を持つ卓上マスコットとして試作した(図5参照)。このマスコットは、筐(きょう)体の外周上に音声認識・音源方向推定用の8個のマイク、合成音声再生用のスピーカ、顔画像認識用のカメラを装備している。さらに、ディスプレイや、旋回が可能な頭部と向きや形状を変更できる耳を装備して、表情などを表現できる。今回試作したうさぎ形マスコットによる研究をさらに進め、将来は、すべての機能をAV機器本体内に組み込むことを検討している。

## 5 おわりに

ここでは、日立製作所が取り組んでいる「人に優しいハイビジョン映像ライフ」の実現に向け、メディア処理技術を有機的に結合した知的ユーザーインターフェースのコンセプトと、これを構成する要素技術について述べた。

今後は、対話インターフェースの自由度を向上させ、さらに多様な操作の実現を目指していく。そのために、日立製作所は、各要素技術のいっそうの性能向上の推進と、インターネットを利用した情報検索により、ホームネットワークを通じて、AV機器だけにとどまらず、情報家電全

体の知的な制御へと、その機能を拡張していく考えである。

### 参考文献

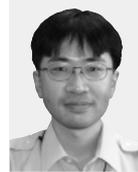
- 1) 戸上, 外: 周波数振り分け法の出力結果に基づく最小分散ビームフォーマの適応化方式, 音学講論(2005.9)
- 2) 小窪, 外: 車載用音声認識における騒音対策とその評価, 電子情報通信学会論文誌, Vol. 83-D, No.11, pp.2190~2197(2000.11)
- 3) Nukaga, et al.: Unit Selection Using Pitch Synchronous Cross Correlation for Japanese Concatenative Speech Synthesis, 5th ISCA Speech Synthesis Workshop(2004)

### 執筆者紹介



#### 新庄 広

1990年日立製作所入社, 中央研究所 知能システム研究部所属  
現在, 顔画像認識の研究に従事  
電子情報通信学会会員, 情報処理学会会員  
E-mail:shinjo@crl.hitachi.co.jp



#### 山口 宇唯

1999年日立製作所入社, 中央研究所 知能システム研究部所属  
現在, 音声対話型ヒューマンマシン インターフェースの研究に従事  
精密工学会会員  
E-mail:ui-yama@crl.hitachi.co.jp



#### 天野 明雄

1981年日立製作所入社, 中央研究所 知能システム研究部所属  
現在, 音声認識技術の研究開発に従事  
電子情報通信学会会員, 人工知能学会会員, 日本音響学会会員  
E-mail:amano@crl.hitachi.co.jp



#### 内部こなぎ

1992年日立製作所入社, 中央研究所 知能システム研究部所属  
現在, ユーザープロフィール応用の研究に従事  
E-mail:konagi@crl.hitachi.co.jp



#### 石橋 厚

1983年日立製作所入社, デザイン本部 ホームソリューションデザイン部 所属  
現在, 主にAV機器のデザインに従事  
E-mail:a-ishibashi@design.hitachi.co.jp



#### 桑本 英樹

1987年日立製作所入社, コヒキタスプラットフォームグループ 事業企画本部 所属  
現在, 同グループの研究開発企画に従事  
電子情報通信学会会員  
E-mail:hideki.kuwamoto.vk@hitachi.com