

K-079

情報家電操作のためのマルチモーダル・インタフェースの可能性 A feasibility study on multimodal interface for handling information appliances

酒井 俊幸†
Toshiyuki Sakai

木原 千怜†
Chisato Kihara

伊田 政樹‡
Masaki Ida

柳田 益造†
Masuzo Yanagida

1. はじめに

ブロードバンドの普及、IPv6、無線LANなど規格の整備により、多種多様な機器を様々な場所でネットワークに接続できるようになった。家庭内では家電がネットワークに接続され、多機能化・高機能化が進み、一般家庭におけるインターネット端末としての役割を求められている。このため、家電には機能選択や文字入力が必要なインタフェースが必要となる。しかし、従来のボタン入力のみで機能をすべて実現した場合、操作が非常に複雑になる。そこで、誰でも簡単に所望の情報を取得できるインタフェースが必要となる [1, 2]。

上記のようなインタフェースとして、我々は音声入力に着目した。音声入力には、誰でも簡単に使用できる点、言葉を用いた概念的な操作ができる点、入力デバイスを小型化できる点、の3つの利点がある。

本研究では音声入力の有効性を評価するためのワークベンチを試作し、応用としてTV操作システムを構築し、情報家電における音声入力の評価・検討を行った。

2. 音声インタフェース評価ワークベンチ

2.1 概要

情報家電のインタフェースの形態として、リモコンに音声入力を付加した端末を想定する。そこで、音声入力端末に以下の2条件を課した。

- 小型・軽量・低コストとする
- 無線端末とする

端末を小型・軽量・低コストにすると、高精度の音声認識を端末上で行う計算リソースを確保できない。そこで、システムをクライアント・サーバ方式とし、音声データの送受信を行う。音声データは容量が大きいため、送受信を行うには十分な帯域が必要である。赤外線通信やBluetoothなどの規格では必要な通信容量を確保できないため、無線LANの広帯域な通信規格を用いる。

2.2 システム構成

構成を図1に示す。本音声インタフェースは、音声入力端末と音声認識サーバによって構成されている。音声入力端末は、PDA上のソフトウェアとして実現しており、音声区間を切り出し、音声データを音声認識サーバに送信する。音声認識サーバでは、送られてきた音声データを、音声認識システムJulian [3]を用いて認識する。この認識結果を基にして情報家電の制御を行う。

3. ネットワーク対応TV操作システム

3.1 概要

前節で作成したワークベンチの応用例として、TVを音声で操作するシステムを作成した。TVは地上波デジ

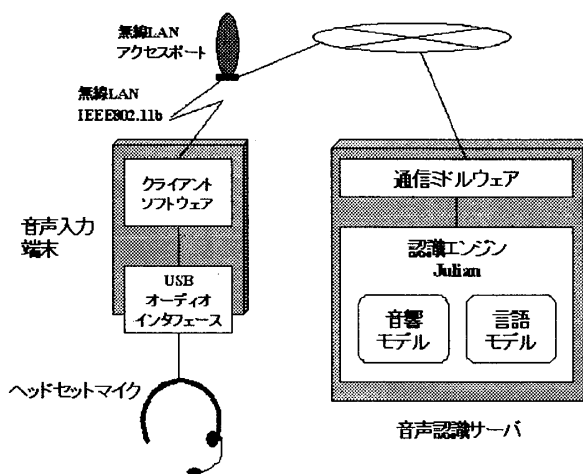


図1: ワークベンチシステムの構成

タル放送の開始により、チャンネル数の増加、番組検索機能やデータ放送による情報配信機能の付加など、多機能化・高機能化が進み、操作が複雑になっている。また、TVは放送される番組が毎日更新されるため、番組情報を用いて操作するためにはあらかじめ放送される番組に対応したコマンド群を用意しておく必要がある。以上の背景から近未来の比較的高度な利用を伴う情報家電として、TVを取り上げ、その操作システムを構築した。

3.2 システム構成

TV操作システムのシステム構成を図2に示す。オンライン番組表から番組データを抽出し、認識辞書と番組情報テーブルを作成する。音声認識サーバからの認識結果をテレビ駆動用コマンドに変換し送出する。

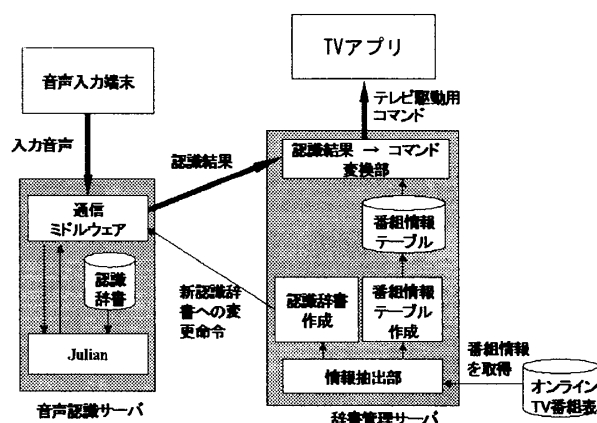


図2: TV操作システムの構成

作成したTV操作システムは、リモコンで操作可能な「電源ON/OFF、ボリューム調整、チャンネル変更」と「番組名・テレビ局名からのチャンネル変更」を音声で

†同志社大学工学部

‡オムロン株式会社技術本部コントロール研究所

操作することができる。番組名の発話によるチャンネル変更は、認識結果と現在の時刻、番組情報テーブルを照合し、現在放送されている番組名である場合はチャンネルを変更する。

4. 評価実験

4.1 実験環境

6畳相当の室内にTV操作システムを設置し、被験者にシステムを使用してもらった。被験者は20~25歳の男女計11名(男性8名,女性3名),発話回数は一人約60発話である。

4.2 評価方法

TV操作システムを実際に使用してもらい、以下の項目についてアンケートを行った。

- TVリモコンの不便なところについてのコメント
- 番組検索が実現すれば音声・TVリモコンのどちらで操作を行いたい
- 本システムの機能のうち音声で操作したい機能
- 本システムに対する要望

また、被験者がTV操作システムを利用する発話内容を記録し、TV操作の成功回数を測定した。

4.3 結果

アンケートの集計の結果の一部を以下に示す。

TVリモコンの不便なところ

- ボタンが多すぎてどのボタンがどの機能に対応するのかわかりにくい
- 画面設定などの詳細な設定をどのようにすれば良いかわからない
- 筐体が大きくて重い

番組検索

- 音声が良い(7名)
 - 画面表示だと自分で探して選択するのに時間がかかる。
 - リモコンで文字入力をする時間がかかる
- TVリモコンが良い(1名)
 - 誤認識を考えるとリモコンの方が確実
- その他(3名)
 - 番組数が多ければ音声, 少なければリモコン
 - 両方使えたら良い

システムへの要望

- 今放送している番組を表示するようにして欲しい
- 入力されたか解り難い
(音声入力がシステムに受理されたかわからない)
- 相槌などの返答が欲しい
- もっと普通に話しても認識して欲しい
- マイクを意識しないでいいようにして欲しい
(ハンズフリー入力の実現)

音声で操作したい機能

- 電源 ON/OFF
- テレビ局名・番組名でのチャンネル変更

操作の成功回数

- 平均 72.3 % (最高 91.8 %, 最低 57.8 %)

以上の結果より、次のことがわかった。

- 音声入力で有効な操作は、画面の詳細設定などの階層構造の深いところにあり時間のかかる操作、電源 ON/OFF のように確定的操作、番組検索のように多くのデータの中から一つのデータを選択する操作、番組検索・テレビ局名・番組名でのチャンネル変更のように文字入力が必要な操作であることがわかった。
- 音声入力が有効でない操作は、ボリューム調整、チャンネル変更など連続して類似動作を繰り返す操作であることがわかった。
- 番組検索の「その他」の回答項目に見られるように、すべてを音声で操作するのではなくリモコンとの併用が必要であることがわかった。今後、音声とリモコンの使い分けについて、ユーザの行動を分析する必要がある。
- 番組検索においてリモコンが優位と回答した被験者は認識率が低い(60.7%)ため、音声操作に対する評価が相対的に低くなったと考えられる。操作の成功回数を見ても、個人によって認識率にばらつきがあるため音声認識性能の改善が必要である。
- 現在放送されている番組名の表示、音声入力に対する応答、ハンズフリー入力の導入などユーザの負担を軽くする機能をシステムに追加する必要がある。

5. まとめ

本稿では、情報家電向けの音声インタフェースとして利用できるワークベンチを設計・試作した。また、ワークベンチをTV操作システムに実装し、被験者による主観的評価を行った。

アンケート結果より、複雑な操作や大量のデータを扱う操作、文字入力を行う必要のある操作に関しては音声入力の方が有効であることがわかった。また、すべてを音声で操作するのではなくリモコンとの併用が必要であることがわかった。

今後の課題として、アンケート結果をもとに音声入力が有効な操作とボタン入力が有効な操作の分類、番組名表示や音声入力に対する応答などシステムからユーザに働きかける機能の追加、音声認識性能の改善、TV以外の家電へ拡張し有効性を検証することが挙げられる。

謝辞 本研究は文部科学省知的クラスター創成事業「ネオカデン」からの援助を受けた。

参考文献

- [1] 美濃他, 情報家電の考え方, 情報処理, Vol.42, No.11, pp.1049-1054, 2001.11
- [2] 土井, ユーザから見た「情報」家電, 情報処理, 情報処理, Vol.42, No.11, pp.1055-1062, 2001.11
- [3] 河原他, 連続音声認識コンソーシアム2002年度ソフトウェアの概要, 情報研報, 2003-SLP-48-1, 2003.10