

G-004

次世代情報家電操作のための音声コマンド設計に関する主観評価実験

Subjective evaluation of speech input aiming at operating next-generation information appliances

木原千怜† 酒井俊幸† 伊田政樹‡ 柳田益造†
Chisato KIHARA Toshiyuki SAKAI Masaki IDA Masuzo YANAGIDA

1. はじめに

情報化社会の発展につれて、情報家電は、多機能化に伴って扱う情報が多様化し、操作がますます複雑になってきている。一方、家電は多くの人が日常的に使用する機器であるため、誰でも簡単に使える必要がある。そこで我々は、情報家電のインタフェースとして音声に注目した。音声は誰でも使えるが、情報家電のインタフェースとして簡単に使えるようにするためのコマンドの設計法や設計指針は確立されていない。そこで情報家電の一例として TV を取り上げ、さまざまなコマンドの使用についてユーザ主観評価実験を実施し、コマンド群の設計の違いが音声インタフェースの有効性に与える影響について検討を行った。

2. ネットワーク TV システム

情報家電の一例として、ネットワーク TV について検討する。TV は近年、使用チャンネル数の増加やネットワークとの融合に伴い多機能化・高機能化が著しい。多機能化・高機能化は操作を複雑にし、現在のリモコンのようなボタンによる操作は、誰でも簡単に使えるインタフェースではなくなる。そこで、音声インタフェースの利用に注目した。音声による入力は特別な道具や学習を必要としない簡単なインタフェースであるといえる。簡単な操作を実現するには、TV システムの作成にあたり、音声コマンドとしてどのような語彙をどのくらい用意しておくことが有用であるか明確にする必要がある。そこで、従来のリモコンボタンの操作を実現するコマンド群のシステムや、リモコンボタンの操作に加え番組名によってチャンネル変更を行うことができるコマンド群のシステムなど、さまざまなコマンド群のシステムを作成し、ユーザの主観評価実験を行う。

筆者らは、音声インタフェースの評価のための TV システムをすでに作成している[1]。システム構成図を図 1 に示す。音声入力端末は、PDA 上にクライアント・ソフトウェアとして実装し、音声の入力には接話マイクを用いている。入力された音声は、PDA クライアントで音声区間が切出され、音声波形がそのまま音声認識サーバに送られる。音声認識には Julian[2]を用いる。情報家電操作のためのコマンドは比較的簡単で短いので、言語モデルは決定性有限状態オートマトンとして記述する。また、音響モデルは IPA(Information Promotion Agency)モデル[2]を用いる。認識結果から、あらかじめ決めておいたコマンド変換用テーブルを用いて TV 駆動用コマンドを作成し、TV を駆動する。

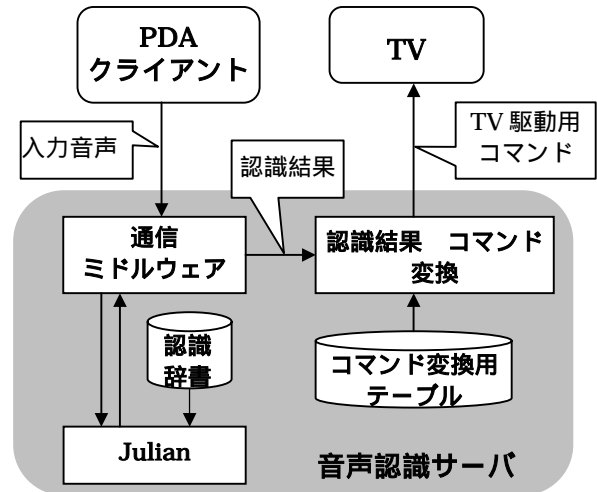


図1. ネットワーク TV システム

3. オンライン番組表からの番組名の取得

ネットワーク TV システムにおいて、音声コマンドとして番組名を用いる場合、番組名は日々入れ替わるため、あらかじめ認識辞書に登録しておくことができない。そこで、Web で公開されているオンライン番組表から番組名などを取得し、認識辞書に登録することで番組名によるチャンネル変更を可能とする TV システムを実現した。認識辞書に単語に登録するには単語の音素表記が必要である。番組名から音素表記への変換は、chasen ver.2.3.3[3]を用いて番組名に読みを付与し、読み仮名から音素への変換は Julian に添付されている yomi2voca.pl を用いた。

4. 評価実験

4種類のコマンド群を用いて 11 人の成人男女にシステムを試用してもらい、それぞれのシステムについて、従来のリモコン操作と比較した使用感の 5 段階評価を集計した。

使用したコマンドの内容と単語数は以下の通りである。

- (1) 現在の TV リモコンの操作を音声で行うためのコマンド (58 単語)
- (2) (1)に放送局名を追加したコマンド (76 単語)
- (3) (2)に番組名を追加したコマンド (423 単語)
- (4) (3)に番組名から考えられる略称や通称を手作業で追加したコマンド (462 単語)

各システムに対して、1~5 それぞれの評価をつけた人数を表 1 に示す。それぞれのシステムで、最も多い人数がつけた評価を網掛けで示す。表 1 より、ユーザにとって使いやすいシステムは、使いやすい順に(4)・(2)・(1)・(3)であることがわかる。

†同志社大学大学院

‡オムロン株式会社技術本部コントロール研究所

表1. ユーザ評価結果

		評価					平均
		使いにくい		使いやすい			
		1	2	3	4	5	
システム	(1)	1	2	7	1	0	2.7
	(2)	0	2	4	4	1	3.4
	(3)	1	6	1	3	0	2.5
	(4)	1	0	2	6	2	3.7

それぞれのシステムに対するコメントの一部を示す。

(1)のシステム

- ◆ ときどき誤認識するのでリモコンの方が確実にチャンネル変更できる(特に数字は誤認識しやすい)
- ◆ リモコンより動作するまでの時間が長いので使いにくい
- ◆ チャンネルをランダムに変えて見たい番組を探すとき、リモコンの方が速くて便利

(2)のシステム

- ◆ 放送局名でもチャンネル変更ができ、チャンネル変更の操作の幅が広がったため便利
- ◆ 放送局名よりチャンネル番号で覚えているため(1)のシステムと操作性は変わらない。チャンネル数が増えると便利かもしれない

(3)のシステム

- ◆ 番組名のコマンドが長すぎて言いにくい
- ◆ 番組名の正式名称を覚えていないので使いやすいと言えない
- ◆ 見たい番組を探すときチャンネル番号を考える必要がなく便利

(4)のシステム

- ◆ 大まかに番組名を覚えていれば良いので便利
- ◆ 短い言葉を言うだけで済むので便利
- ◆ 自分の考える略称が認識語彙にないので不便

また、誤認識しない理想的な音声認識が実現したと仮定したとき、次に示す操作のうち、それぞれ「音声操作が有効な操作である」・「ボタン操作が有効な操作である」と答えた人数を表2に示す。

- ・電源の ON/OFF
- ・チャンネル番号によるチャンネル変更
- ・ボリューム調整
- ・テレビ局名によるチャンネル変更
- ・番組名によるチャンネル変更

表2. 操作に対する音声の有効性

	音声操作 が有効	ボタン操作 が有効	どちらとも 言えない
電源 ON/OFF	7	2	2
チャンネル番号	4	5	2
ボリューム調整	3	6	2
テレビ局名	7	3	1
番組名	8	2	1

以上の実験により、次のことが分かった。

- テレビ局名や番組名は入力する文字数が多いため、音声による入力が有効である。ただし、番組名の正式名称を必要とする場合には、正式名称を覚えていなかったり、長いコマンドは発声しにくかったりする問題がある。この問題点は略称や通称を音声コマンドに追加することで解決できる。
- ボリュームの調節やチャンネルをランダムに変えたい時のように、短時間内に同じ操作を何度も行う場合には音声による操作が使いにくいと感じる。ただし、電源の ON/OFF のように一度だけの操作の場合には、単純な操作でも便利であると感じられる。
- 現時点での TV 操作では放送局名を意識しないユーザもいる。しかし、チャンネル数が増えた時、チャンネル番号よりも放送局名を覚えてチャンネル変更を行うようになるならば、放送局名を音声コマンドに追加することは効果的であると言える。
- 音声コマンドはユーザにとって直感的に分かりやすいものである必要がある。直感的に分かりやすい音声コマンドをシステムに追加した場合、ユーザは使用できるコマンドを考えることなくシステムを使えるため、使いやすいシステムになると言える。
- ユーザにとって使いやすい音声コマンドを用いれば、音声認識システムは多少の誤認識があっても使いやすいと感じる。

5. まとめ

本稿では、情報家電に対する音声インタフェースの利用について、音声コマンドの設計法や設計指針の検討を行った。情報家電の操作のように複雑な操作に対し、音声を利用することの効果を確認できた。ただし、音声コマンドは短いものでなければならず、直感的に分かるものでなければならない。すなわち、音声コマンドとしてユーザの感覚に合った略称・通称を追加することが音声インタフェースの使いやすさを高めるといえる。

しかしながら、すべてのユーザの感覚に合った略称・通称を網羅的に書くことは難しい。そこで、今後の課題として、どのような略称・通称がより多くのユーザの知識に合っているのか、またどのようにその略称・通称を取得するのか検討が必要である。

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] 伊田他, 次世代情報家電制御のための音声インタフェースの試作, 情報処理学会全国大会, 2004.3
- [2] 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, 情処研報, 2003-SLP-48-1, 2003.10
- [3] 松本他, 日本語形態素解析システム『茶釜』使用説明書, 2000.12