

音声認識と音声合成を用いた ビデオ機器の遠隔操作に関する検討

加藤 誠巳 田平 武彦
(上智大学理工学部)

6N-9

1 まえがき

近年、一般家庭にもPCが普及しつつある。このPCの性能の飛躍的な向上により、音声認識や音声合成が特別なハードウェアを用いることなくソフトウェアで可能となって来た。本稿では、PCによる音声認識及び音声合成を用いて、ユーザにとって面倒と考えられる家庭電化製品の操作を簡単化する方法について検討を行い、さらに、認識辞書の切り替えによって、認識率の向上を図った結果について報告する。

2 システムの概要

ここでは、初心者には操作が難しいと思われる家庭電化製品の例として、VTRの録画予約を対象とした。現時点ではPCと通常のVTRとの間には接続インターフェイスがないので、主な動作はPC画面上の仮想VTRを用いて行うことにした。

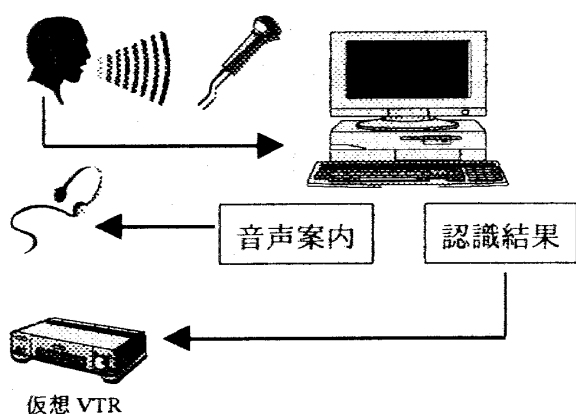


図1 システムの概要

A Remote Control System for VTR by Means of Speech Recognition and Synthesized Voice
Masami KATO, Takehiko TAHIRA
Sophia University

音声認識及び音声合成は Windows95 上で動作する Microsoft の MS-Voice (Speech API)を用い、その上に NEC VOICE を載せ、日本語化している。

本システムは以下のような機能を有している。

- 音声によって VTR の操作を行う
「電源 ON」、「電源 OFF」、「録画」等 VTR に対する一般的な命令用語をそのまま認識することができる。
- VTR の基本操作に特化した辞書をもつ
誤認識を防ぐために、操作に必要な基本的な命令のみを辞書に収録した。
- 入力可能な命令を合成音声を用いてユーザに伝える
現在どのような命令が入力可能であるのかを、音声と画面上のウィンドウによってユーザに通知し、辞書に存在しない単語の入力を減らし誤認識を防いでいる。
- 同義異音語の収録
テレビ局名等は人によって、いろいろな呼び方をする。たとえば「日本テレビ」などは、「日テレ」、「4チャンネル」など、複数の単語が同意語として用いられている。これらを全て辞書に収録し、不特定の人による操作を可能としている。
- 認識用辞書の切り替えによる認識率の向上
辞書の収録単語数の増加は、認識率の低下につながる。一方、多数の同意語の辞書への収録は操作性の向上につながる。そのため、頻繁に使用する辞書を切り替え、認識率の向上と操作性の向上の両立を図っている。
- NEC VOICE の認識用辞書のための工夫
NEC VOICE では、認識させる単語をひらがなの読みによって登録するようになっている。しかし、

アルファベットの読みなどは人によって異なるので、以下のような登録を行っている。

対象とする単語	読み
NHK	えぬえいちけー
NHK	えぬえちけー
NHK	えぬえっちけー
NHK	えぬえいちけい
NHK	えぬえちけい
NHK	えぬえっちけい

図2 アルファベットの読みの例

3 認識率の測定

辞書を分割した場合と、そうでない場合の認識率の実験を行った。実験の方法は以下に示す通りである。

1. 「テレビ局（東京の7つのVHF局）名」、「時間」、「3倍・標準」の3分野に分けた辞書と、認識すべき単語全てを含む辞書を用意する。
2. 任意の局名、任意の時間帯、3倍・標準を1つの単位とする。
例) NHK 朝6時から朝8時 3倍
3. NEC VOICE 添付のユーザ学習を5回行い、10単位を1組とし、分割辞書、全ての単語を含む辞書両者に対して、音声による入力を行う。このとき、テレビ局名や時刻は、辞書中の表現を全て網羅した。
4. 1単位中に、チャンネル、開始時間、終了時間、標準または3倍、と4つの認識すべき単語があるので、1組40単語の内、いくつ正しく認識されたかを認識率とし、5組の結果を平均して算出した。

認識率の結果を右上に示す。辞書をこまめに切り替えることによって、明らかに認識率が向上した。全ての単語を一つの辞書に入れた場合の認識率の低下は、主に短い単語での誤認識が多くなるのが原因である。適切な分野ごとに辞書を複数作成し、必要に応じて切り替えを行うことによって、認識率は大幅に向上し、実用に耐えるレベルの認識性能を得ることが可能となる。

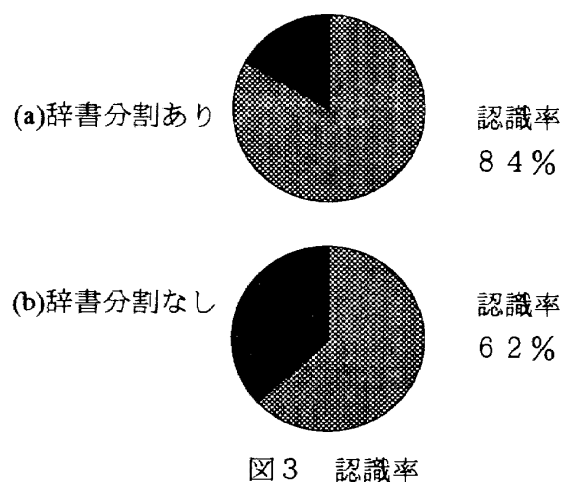


図3 認識率

4 むすび

マニュアルの操作手順に従って、録画の予約を行うことは、初心者ユーザにとってはかなり面倒な作業であるが、音声を仲介とすることによって、メカに弱い人にも簡単で、親しみやすいユーザインタフェースを構築することが可能となる。また、音声認識では認識率が高いほど操作性は向上するが、認識辞書の適切な収録分けと切り替えを行うことにより、さらに認識率を向上させることも可能と考えられる。

今後、家庭電化製品にコントロール IrDA 規格を持つものが登場してくると思われる。現在は、PC による家電製品の操作・制御はあまり一般的ではないが、この規格を用いることによって、PC による家電製品のさらに高度な操作が可能になると考えられ、PC を応用した新たなユーザーインタフェースが簡単に実現可能となると考えられる。

最後に、有益な御討論を戴いた本学マルチメディア・ラボの諸氏に謝意を表す。

参考文献

- [1]加藤,田平:“音声に意識を用いて電話回線を介した IrDA による機器の遠隔操作システム”,情処学会第55回全大,5L-7(平成9年9月).