

A-19-5

複数の認識エンジンを併用したビデオ操作支援システム

Video operation support system that used recognition engine together

小倉英樹†
Hideki Ogura

片山憲昭†
Noriaki Katayama

島田伸敬†
Nobutaka Shimada

白井良明†
Yoshiaki Shirai

立命館大学情報理工学部知能情報学科†

1 はじめに

近年音声対話システムの構築が盛んに行われている。単純な質問、応答を行うだけでなく、対話によりユーザと情報の交換をし、ユーザが発話する様々な文章（短文、長文、同義の文）を認識し、意図を正しく理解できる対話システムの構築を目的としている [1]。また、音声認識結果が誤った時やユーザが会話の流れから予期しない発話をした時も、必要に応じて適切に質問や確認することで、機能を正しく実行する対話システムを構築する。

本研究では、上に挙げた対話システムを作る際に、長文の認識を得意とする Julius と単語や短い発話を正しく認識できる Julian を組み合わせてユーザの様々な発話に柔軟に対応できる対話システムを構築する。

2 システム構成

本研究に使用した機器は、無線 LAN 内蔵のノート型 PC、ネットワークサーバ機能を持つ HDD レコーダ、マイクとスピーカである。

そして図 1 に示すように、本システムは対話部、制御部そして通信部と分けられる。対話部ではユーザの発話を認識し制御部へ認識結果を渡し、また制御部から渡された質問文をユーザに対してシステム発話する。制御部ではユーザ発話の認識結果を処理し、質問文を生成する。そしてシステムの必要に応じて通信部で HDD レコーダと通信しビデオ操作、あるいは情報を取得するといったシステムである。

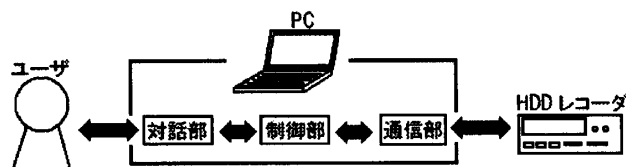


図1 システム概要

3 Julius による認識の問題点

[1]の研究では、音声認識に Julius のみを用いている。そのため、ユーザ発話が長文の場合はうまく認識を行うが、一単語程度の発話では認識を誤ることが度々起こる。例えば、ユーザ発話が「1時（いちじ）」と発話したのに対し、読みの似ている「7時（しちじ）」と誤認識をすることがある。

そこで、あらかじめユーザ発話が1単語程度と予想される質問を行った場合、短文の認識を得意とする Julian にも同時に認識させる。そうすることで、長い文章でも短い文章でもそれぞれ得意とする認識エンジンによって認識を行うことができる。

4 Julius, Julian の並列使用

4.1 Julian 認識に必要な辞書作成

ビデオ操作のうち予約録画機能における対話について説明する。今回作成した辞書の例を次に示す。1つ目は0時

から12時まで時間を記述した辞書 (hour)。2つ目はチャンネル情報を記述した辞書 (channel)。最後に Julian に認識させたくない際に何も書かれていない辞書 (blank) を作成した。

4.2 対話中の Julian 辞書の差し替え

本研究では、必要に応じて Julian の認識用の辞書を動的に差し替えて使用する。システム起動時は辞書 blank を設定しておく。対話の中で開始・終了時刻そしてチャンネル情報が欲しい際に、システム発話すると同時に辞書 hour または channel を設定する。そしてユーザ発話が行われると、Julian に設定した辞書から認識が行われる。また、話題転換などにより Julian で認識できなかった時は Julius の認識結果を用いる。最後に認識辞書を blank あるいはシステム発話に基づいて変更する。

5 評価実験

図2に従来システム [1]の結果を示す。「しちじ」と発話したのに対し、読みの似ている1時として誤認識している。図3は Julian に辞書を設定した際に、Julian のみで認識を行った結果である。長文のユーザ発話は認識できない。図4に本研究で提案したシステムの結果を示す。上記の問題点を改善し、正しい認識と処理を行うことができた。

```
User : 今日6時から帰って
System : 今日6時からですね。終了時刻は何時ですか
User : 1時(に誤認識)
System : 終了時刻を開始時刻より前に指定して下さい
```

図2 認識エンジン:Julius

```
User : 今日6時から帰って
System : 今日6時からですね。終了時刻は何時ですか -----> 辞書 hour に設定
User : 7時
System : 7時までですね。何チャンネル録画しますか -----> 辞書 channel に設定
User : やっぱり5時から帰ってください
System : すいません。もう一度言ってください -----> 辞書 blank に設定
```

図3 認識エンジン:Julian

```
User : 今日6時から帰って
System : 今日6時からですね。終了時刻は何時ですか -----> 辞書 hour に設定
User : 7時
System : 7時までですね。何チャンネル録画しますか -----> 辞書 channel に設定
User : やっぱり5時から帰ってください
System : 今日6時から7時までですね。何チャンネルですか。 -----> 辞書 blank に設定
```

図4 認識エンジン:Julius, Julian

6 まとめ

本稿では Julius と Julian を使用し、長文・短文の両方を認識する手法を提案した。誤認識は減り、Julian も使用することで単語程度の認識は確実に良くなった。

7 関連、参考文献

[1] 片山憲昭, 島田伸敬, "音声によるビデオ操作支援システム", 電気電子情報通信学会総合大会, 2005.3