

機関番号：32689

研究種目：若手研究(B)

研究期間：2009～2010

課題番号：21700204

研究課題名(和文) 音声対話システム応用のための音声認識に関する研究

研究課題名(英文) Study on Speech Recognition for Spoken Dialog System

研究代表者

藤江 真也 (FUJIE SHINYA)

早稲田大学・高等研究所・助教

研究者番号：00367062

研究成果の概要(和文)：

従来の音声認識システムに不足していた、音声対話システムに応用するための機能を実現した。具体的には、入力音声断片化している場合に複数の断片を統合して認識する機能、音声認識結果と発話音声の持つ韻律情報から発話終了を推定する機能を実現した。これらの機能を応用して、適切なタイミングで発話を開始することのできる音声対話システムを実現し、会話ロボット上に実装した。

研究成果の概要(英文)：

Missing functions of a speech recognition system for a spoken dialog system has been developed. Especially, recognition with integration of speech segments of fragmented utterance, and utterance end detection with speech recognition results and prosodic information have been implemented. These functions are utilized to develop a spoken dialog system that can take its turn at appropriate timing on a conversational robot.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
21年度	1,400,000	420,000	1,820,000
22年度	1,200,000	360,000	1,560,000
年度			
年度			
年度			
総計	2,600,000	780,000	3,380,000

研究分野：総合領域

科研費の分科・細目：知覚情報処理・知能ロボティクス

キーワード：音声情報処理，音声認識，音声対話システム

## 1. 研究開始当初の背景

音声を用いた対話は人同士で日常的に行われる。人にとって標準的とも言えるこの手段を用いたインターフェースの実現が望まれ始めて久しいが、未だ実用的な音声対話システムが構築された例は無い。従来の音声認識システムは、音声対話システムに応用するという観点から見た場合に以下のような問題がある。

a. 文(あるいは単語列)が認識単位であり、

文頭から文末までが一息のうちに発声されることが前提となっている。

b. 認識対象となる文が事前に決められ、システム実行後は変更することが出来ない。

c. 対話で必要となる他の情報(発話態度や発話番制御情報)について考慮されていない。

これらの問題を解決することによって、音声対話システムに適した音声認識システムを

構築すること、またそれを用いることで、より自然な対話が可能な音声対話システムを構築することが可能となる。

## 2. 研究の目的

音声対話システムに応用するという観点から見た場合に不足している音声認識システムの機能を検討し、実際の音声認識システムに適用する。具体的には、下記の機能を実現することを目的とする。

- (1) 複数の発話を統合して認識する機能
- (2) 複数の認識対象を保持し、それらを同時に認識する機能
- (3) 発話態度認識等の対話に必要な情報を認識する機能

更に、これらの機能を用いた音声対話システムを構築し、従来の音声対話システムが見逃していたターンテイキングの問題について深く掘り下げることを目的とする。

## 3. 研究の方法

まずはじめに、既存の音声認識システム、音声対話システムが持つ問題点について、実際にそれらのシステムを用いることで検討する。そこで得られた問題点を元に、新しい音声対話システムのデザイン、及びそこで必要となる音声認識システムの機能を決定する。次にそこでデザインされた音声認識システムの構築を行う。最後に、これらを統合した音声対話システムを構築し、会話ロボットに統合した上でデモンストレーションを行う。

## 4. 研究成果

### (1) 会話ロボットによる予備実験

まず、既存の音声認識システムの問題点を洗い出すために、会話ロボット ROBISUKE による会話実験を行い、そこで起こる問題について検討した。

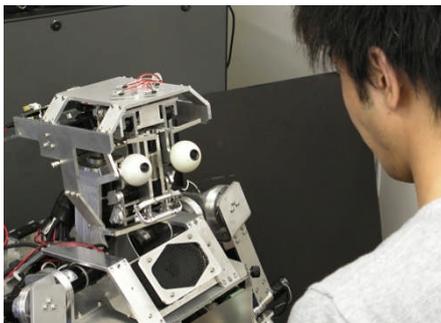


図1 会話ロボット ROBISUKE

結果から、音声認識システムの誤りの原因と

して、人の発話する音声断片化することで文として成立しないことが頻繁に起こること、場面の移り変わりによって認識対象となる文が変化すること、システム側の発話タイミングの問題で人の発話を中断してしまうこと等が挙げられた。これらの問題に対して、音声認識システムを拡張することで解決することを試みた。

### (2) 発話断片化に対応した音声認識システム

一つ目の問題として、音声対話システムに話しかける場合の人の発話が断片化する問題が挙げられた。断片化が起こると、一つの文に対応する音声が入力されることが期待される従来の音声認識システムでは認識精度が著しく劣化することが分かっている。そこで、複数の断片をまとめて一つの入力としてみならず音声認識器を構築した。断片のまとめ方に関しては、それまでの音声認識結果から得られる文末らしさと、ポーズ長の情報を用いることとした。言語モデルには有限状態トランスデューサを用いた。

構築した音声認識器を、N-gram 言語モデルを用いた従来の音声認識器と性能を比較した。音声区間検出のパラメタを変えることで音声認識器の性能は異なるが、どのパラメタにおいても従来の音声認識器よりも優れた結果を示した(図2)。

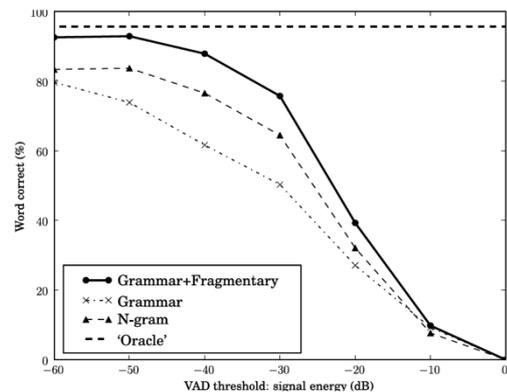


図2. 断片化に対応した音声認識の実験結果

### (3) 音声対話システムの構築

予備実験より明らかになった問題点の一つとして、音声対話システムが発話を開始するタイミングの決定を誤ると、人の発話を妨害してしまうことが挙げられた。この問題を解決するために、音声対話システムの再設計を行った。

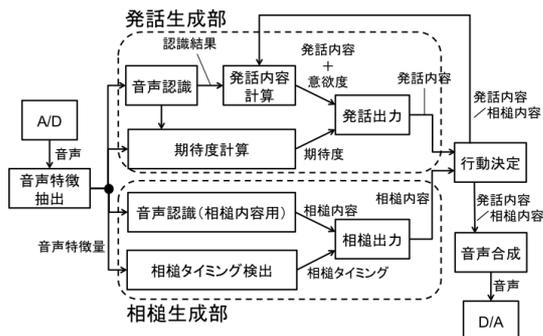


図 3. 音声対話システムの設計

音声システムのモジュール構成図を図 3 に示す。構成は大きく分けて発話生成部と相槌生成部、行動決定部に分かれる。

発話生成部では、ユーザの発話を入力とした音声認識器を実行し、発話断片が入力される毎にその時点でシステムが発話すべき内容を決定する。その時に、その発話がどれほど重要であるかどうかということをもとにシステムが発話をどれくらいしたいか（発話意欲度）を計算する。また、音声認識結果と断片末の韻律情報を基に、ユーザの発話を終了したい（システムに発話権を譲りたい）度合い（発話期待度）を計算する。発話意欲度と発話期待度のバランスから実際に発話を出力するかどうかを決定する。

相槌生成部では、ユーザの発話に割り込む必要の無い短い反応を生成する。これは研究者の従来研究で構築されたものを応用する。

行動決定部では、発話生成部、相槌生成部から出力された発話・相槌内容を音声合成によって合成する。発話と相槌の生成はそれぞれ関連無く非同期で行われるため、システムが発話を行っている最中に相槌を抑制したり、または相槌を行っている最中に生成された発話をバッファリングしたりする役目を果たす。

#### (4) 会話ロボットへの実装

構築した音声対話システムを会話ロボット SCHEMA (図 4) に実装した。このロボットは目やまぶたを持ち、表情生成が可能で、体の向きや腕によるポインティングが可能である。本ロボット上へ、(3) で述べた音声対話システムを実現した。具体的なタスクとしては昼食をとる場所を相談する内容の対話システムを実現した。ユーザが明確な希望を持ってない場合においても、適切に発話を割り込ませながら条件にあった店を提案することを実現した。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)



図 4. 会話ロボット SCHEMA

[雑誌論文] (計 2 件)

- ① 藤江真也, 小林哲則, 基礎講座 第 3 回「音声・音響インタフェース」音声対話システム, ヒューマンインタフェース学会誌, 査読無, vol. 12, no. 3, 2010, pp. 195-200
- ② 藤江真也, 小川哲司, 小林哲則, 会話ロボットとその聴覚機能, 日本ロボット学会誌, 査読無, vol. 28, no. 1, 2010, pp. 23-26

[学会発表] (計 1 2 件)

- ① Y. Matsuyama, S. Fujie, H. Taniyama, and T. Kobayashi, “Framework of Communication Activation Robot Participating in Multiparty Conversation,” AAAI2010 Fall Symposia, Arlington, VA, Nov. 11, 2010
- ② S. Fujie, “Conversational Robot and Its Application,” Young Researchers’ Round Table on Spoken Dialogue Systems, Sept. 23, Tokyo, Japan, 2010.
- ③ Y. Matsuyama, S. Fujie, H. Taniyama and T. Kobayashi, “Psychological Evaluation of A Group Communication Activation Robot in A Party Game,” Proc. Interspeech2010, pp. 3046-3049, Makuhari, Japan, Sept. 30, 2010.
- ④ S. Fujie, Y. Matsuyama, H. Taniyama, and T. Kobayashi, “Conversation robot participating in and activating a group communication,” Proc. Interspeech2009, pp. 264-267, Brighton, UK, Sept. 9, 2009.

- ⑤ 藤江真也, “多感覚情報を利用した音声会話ロボット,” 認知科学会 2010 冬のシンポジウム, 東京, Dec. 11, 2010.
- ⑥ 藤江真也, 松山洋一, 小林哲則, “グループコミュニケーション活性化ロボット,” 人工知能学会 言語・音声理解と対話処理研究会, SIG-SLUD-B002, pp.7-10, 東京, Oct. 28, 2010.
- ⑦ 松山洋一, 藤江真也, 齋藤彰弘, Xu Yushi, 小林哲則, “コミュニケーション活性化を指向した会話ロボット ～通所介護施設における事例,” 信学技報 vol.10, no.219, pp.7-12, 千葉, Oct. 8, 2010.
- ⑧ 藤江真也, 小林哲則, “Haar-like 特徴量を用いたカスケード接続型識別器による音声単語検出,” 日本音響学会秋季研究発表会, pp.45-48, 大阪, Sept. 15, 2010.
- ⑨ 松山洋一, 谷山 輝, 藤江真也, 小林哲則, “グループコミュニケーション活性化ロボットの印象評価,” 人工知能学会 言語・音声理解と対話処理研究会, SIG-SLUD-B001, 長崎, pp.7-12, July 22, 2010.
- ⑩ 谷山 輝, 松山洋一, 藤江真也, 小林哲則, “参与構造を考慮した行動設計に基づくグループ会話ロボットの構築,” 人工知能学会 言語・音声理解と対話処理研究会, SIG-SLUD-A903, pp.55-60, 兵庫, Feb. 13, 2010.
- ⑪ 藤江真也, “会話ロボット研究の現状と課題,” 人工知能学会 言語・音声理解と対話処理研究会, SIG-SLUD-A902, pp.29-34, 東京, Nov. 20, 2009.
- ⑫ 谷口 徹, 藤江真也, 小林哲則, “発話継続の逐次的判断機能を備えた音声認識システム,” 日本音響学会秋季研究発表会, pp.221-224, 福島, Sept. 17, 2009.

[図書] (計 1 件)

- ① “音声言語処理の潮流,” 第 5 章 音声対話システム, 白井克彦 編著, pp.175-204, コロナ社, March 2010.

## 6. 研究組織

### (1) 研究代表者

藤江 真也 (FUJIE SHINYA)  
早稲田大学・高等研究所・助教  
研究者番号: 00367062