

状況検知を利用したロボット用音声認識インタフェースの一手法

A Speech Recognition Interface for Robots using notification of ill-suited conditions.

F-11

岩沢 透

Toru IWASAWA

大中 慎一

Shin'ichi OHNAKA

藤田 善弘

Yoshihiro FUJITA

1 はじめに

近年、音声認識を利用し人と対話することを目的としたロボットの研究開発が実用化を目指し進められている[1][2]. 我々はパーソナルロボットPaPeRoの研究開発において、家庭環境での利用を想定した離散単語音声認識インタフェースの研究開発を行っている[3]. PaPeRoの目指す音声認識インタフェースは、ハンズフリーでかつ発話距離に融通性を持つ、雑音に対して頑健で利用者の身体的自由度の高い音声認識インタフェースである.

このような実環境での使用を前提とした音声認識インタフェースの構築には様々な音声認識精度の劣化を引き起こす問題がある. これらの問題は発声不良や未知語発話といった利用者側の発話に関する問題とシステム内外から混入される雑音に起因する環境の問題に大別される. しかしながら、現状の音声認識では、これらの利用者側の発話の問題や雑音混入の問題をカバーし正確な認識結果を得ることは極めて難しい. また、環境面の問題に対しては、マイクロフォンアレイを利用し雑音を除去する技術が研究されているが、実環境において発話音声に重畳される未知で非定常な雑音を除去することはやはり難しい[4].

そこで、我々は音声認識の精度劣化につながる問題状況を検知し、その検知結果を利用して利用者システム双方の適応を図ることで問題状況を回避する状況検知の手法を検討している. 状況検知結果の利用方法としては、利用者に問題状況の改善を依頼するPassiveな応答動作と、システム側で動的に問題状況に対処するActiveな応答動作が考えられるが、今回は警告ガイダンスを利用したPassiveな応答動作をメインにPaPeRoへ実装し評価を行った. 本稿では、状況検知を利用した音声認識インタフェースとPaPeRoへ実装した状況検知の評価について述べる.

2 状況検知を利用した音声認識インタフェースとパーソナルロボット PaPeRo への実装

状況検知は、音声認識時にマイクから入力される音声を解析し音声認識精度の劣化につながる問題状況を抽出することにより行われる. 図1に状況検知を

利用した音声認識システム全体の動作の流れを示す. 状況検知部は入力音声データに対する状況検知結果を返す以下の3つのコンポーネントからなる.

1. 周囲環境の雑音推定を行う環境推定コンポーネント
2. 人間の発話を検出する発話検知コンポーネント
3. 発話の音量やタイミングの問題を検出する発話特性分析コンポーネント

状況検知結果は音声認識エンジンが返す音声認識結果と共に動作選択部で照合され、応答動作が選択される. 音声認識結果は、音声認識辞書の語彙カリジェクトのいずれかが返すものとする.

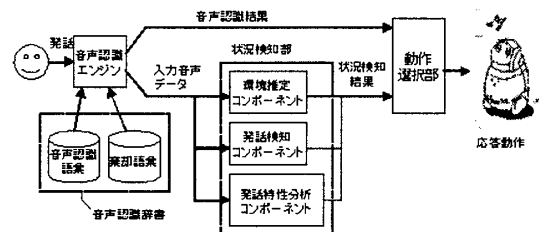


図1: 状況検知を利用した音声認識システムの動作図

PaPeRoへの実装にあたっては、入力音声データに認識エンジンの切り出し音声を利用した. 音声認識辞書には、PaPeRoの認識語彙に加え、文献[3]に記載の棄却辞書を搭載し棄却辞書へのヒットをリジェクトとした. 状況検知部には、環境推定コンポーネントに主に継続雑音を検出する周囲雑音検知を、発話検知コンポーネントに入力音声の調波構造を利用した発話検知を、発話特性分析コンポーネントに入力音声開始直後の発話音声検出により発話タイミングの早すぎを検出する頭切れ検知と音量の問題を検知するパワー不足検知とパワー過多検知を実装した. なお、音量については問題あり、多少問題ありの2段階の閾値を与えそれぞれパワー不足(過多)、パワー不足(過多)気味のように分類し検知する.

次に音声認識結果と状況検知結果の結果に基づくPaPeRoの動作を表1に示す. 音量の状況検知に関しては、2段階の閾値に応じた検知項目別に音声認識結果に対する適用範囲を変えている. ガイダンスの例は、周囲雑音警告が「周りがるさい」「あまりいろいろなこと言わないで」、頭切れ警告が「耳が光っているときに話して*1」、パワー不足(過多)警告が「もう少し大きい(小さい)声で話して」「もっと近

*1 PaPeRo は耳が光っている時のみ音声認識を行う

音声認識結果	状況検知結果	対応動作
-	周囲雑音検知	継続雑音警告
	非発話	無視
	パワー不足	無視
	パワー過多	パワー過多警告
認識語彙	通常発話	通常動作
	頭切れ	頭切れ警告
	パワー不足気味	通常動作
	パワー過多気味	通常動作
リジェクト	通常発話	通常棄却動作
	頭切れ	頭切れ警告
	パワー不足気味	パワー不足警告
	パワー過多気味	パワー過多警告

表1: 動作選択表

くで(離れて)話して」などである。また通常棄却動作は「うんうん」「えっ？」などの一般的な棄却動作を行う。

3 評価

PaPeRoの実環境評価の一環として、実際の家庭環境でPaPeRoを動作させられる入力音声の収集を行っている。ここでは、PaPeRoが警告ガイダンスを行う特定状況検知(周囲雑音、頭切れ、パワー不足(気味)、パワー過多(気味))に関して評価した結果について述べる。評価は、PaPeRoを家庭環境で使用し収集した音声データを利用し経過日数に対する状況検知への適応度合及び音声認識精度に対する寄与について行った。

使用日数の経過に伴う状況検知フィードバックの発生頻度の変化に関して、使用日毎の話しかけ総数に対する発生比率を初日を100として正規化したグラフを図2に示す。図2より、パワー不足警告を除けば、使用日数の経過と共に状況検知の頻度が減ってきており、状況検知のフィードバックに利用者が適応していることが分かる。また、パワー不足の音声は怠け気味の発話が多く、使用日数の経過と共に減少しない原因は利用者の慣れが大きな原因と見られる。パワー不足警告は、このような利用者の慣れに伴う怠け発話を定期的に矯正する役割を持っていると考えられる。

次に、状況検知直後の話しかけ音声(未知語発話は除く)に対する認識率を通常棄却との差として比較した結果を表2に示す。通常棄却と特定状況検知との比較では、雑音検知と頭切れがやや低下、パワー不足とパワー過多が特定状況検知より高い認識率が得られた。特にパワー不足は、検知直後の発話に対する認識率が最も高く利用者の慣れに伴う怠け発話矯正の効果が現れていると考えられる。一方、通常棄却より認識率が劣化した項目に関しては、頭切れは図2より検知が(利用者の不慣れな)使用開始直後に集中している点が認識率低下の原因と考えられるが、周囲雑音検知に関しては利用者が周囲環境へ改善を働きかけることの難しさが現れていると考えられ

検知対象	通常棄却	特定状況検知			
		周囲雑	頭切	パ不足	パ過多
認識率差分	0.0	-3.7	-5.3	+10.9	+5.6

表2: 状況検知フィードバック直後の発話に対する認識率差分。周囲会話などの雑音に関しては、やはりPaPeRo側で何らかの適応を行うことが望まれる。逆に利用者側で適応が容易な発声パワーに関する警告は、検知直後の発話に対する認識率が高く怠け発話の矯正という点も含め認識精度の向上に寄与しているものと考えられる。

また、通常棄却より特定状況検知直後の認識率が高いものが存在する点に関しては、誤認識原因をピンポイントにフィードバックすることが利用者へのシステム状態の透過性を高める効果を持つことが一つの大きな要因と推測される。

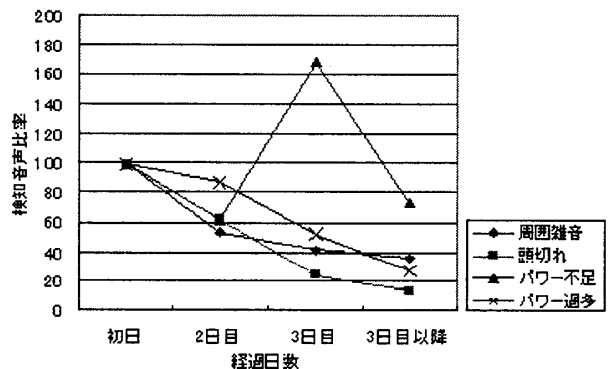


図2: 使用経過日数と状況検知フィードバックの比率

4 おわりに

本稿では、状況検知を利用した音声認識インタフェースの一手法とパーソナルロボットへの適用した評価結果について述べた。今回PaPeRoに実装した警告ガイダンスを行う状況検知では、使用時間の経過に伴う問題状況検知の頻度減少や特に発声パワーに対する警告時の検知後発話に対する認識精度改善の効果が見られた。

今後は、状況検知の精度向上を図ると共に例えばパワー不足検知時にロボット側で人に近づき問題状況の改善を図るなどのActiveな適応についてさらに研究を進める予定である。

参考文献

- [1] 松井, 麻生, J.Fly, 浅野, 本村, 原, 栗田, 速水, 山崎, “オフィス移動ロボット Jijo-2 の音声対話システム”, 日本ロボット学会誌 Vol.18 No.2, pp.300-307, (2000).
- [2] 松坂, 東条, 小林, “グループ会話に参加する対話ロボットの構築”, 電子情報通信学会論文誌 Vol.J84-D-II, No.6, pp.898-908 (2001).
- [3] 岩沢, “パーソナルロボット PaPeRo の音声認識インタフェース”, 人工知能学会 AI チャレンジ研究会論文集, Vol.13, pp.17-23, (2001).
- [4] 中村, “外乱に強い音声認識を目指して”, 日本音響学会誌, Vol.57, No.10, pp.662-667, (2001).