

ロボットの対話と行動の統合モデル

(株)ホンダ・リサーチ・インスティテュート・
ジャパン
中野 幹生

講演の流れ

- ロボットの音声インタフェース
- 音声対話インタフェース研究
- ロボットの行動と対話の統合モデル
- ロボット対話の今後の課題

ロボットの音声インタフェース

ロボット技術の発展

- 近年のロボット技術の発展は目覚ましい
(瀬名, 2001; 日経トレンディ, 2005)
 - 特にヒューマノイドロボット
- ロボット産業・工学における日本のプレゼンスは非常に大きい
 - 産業用ロボット生産台数は世界一

ロボット研究プロジェクト

- 人間協調・共存型ロボットシステム (Humanoid Robotics System)
 - 1998-2002
<http://www.mstc.or.jp/hrp/main.html>
 - ヒューマノイドロボットHRPを開発
- NEDOロボットの開発基盤となるソフトウェア上の基盤整備
 - 平成14～16年
<http://www.nedo.go.jp/activities/portal/gaiyou/p02018.html>
 - RTミドルウェアの開発
- NEDO次世代ロボット実用化プロジェクト
 - 平成16～17年度
<http://www.nedo.go.jp/activities/portal/gaiyou/p04003.html>
 - 愛・地球博での実証実験
- ネットワークロボットフォーラム
 - <http://www.scat.or.jp/nrf/>
- ゆかりプロジェクト
 - http://www2.nict.go.jp/jt/a135/research/ukari_project.html

様々なロボット

- 用途
 - 産業用ロボット
 - パーソナルロボット
 - サービスロボット
- 自律性
 - 自律型
 - 操縦型
- 形状
 - ヒューマノイド(2足歩行, 車輪)
 - 非ヒューマノイド

(厳密な分類ではない)

ロボット音声対話インターフェース

- ヒューマノイドロボットは汎用機械

(長田 2005, 梅谷2005)

→ (音声)コミュニケーション機能が重要

- ほとんどのヒューマノイドロボットは音声インタフェースを持つ

- 音声インターフェースの目的の違い

- サービスロボットへのタスク依頼
- コミュニケーションロボット
- 上記二つの両方を持つもの

Copyright © 2005 Honda Research Institute Japan Co., Ltd.

7

サービスロボットの音声インタフェース

- 人のコマンドを理解
- 小語彙の単語音声認識をベースにしたものが多い
- シナリオに沿っていれば少し複雑な対話も可能

Copyright © 2005 Honda Research Institute Japan Co., Ltd.

8

コミュニケーションロボット

■ コミュニケーションロボット

- 音声言語やジェスチャーを用いてコミュニケーション
- 人がロボットとどうコミュニケーションするかを調べるのにも使われる
 - AIBO, QRIO (Sony)
 - Robovie (ATR)
 - Infanoid (NICT)
 - InterRobot (岡山県立大)
- 必ずしも言語に拘らない
 - Leonardo, Kismet (MIT)

Copyright © 2005 Honda Research Institute Japan Co., Ltd.

9

ロボット対話機能の高度化に向けて

- 音声認識の向上
 - ロボットマイクでの認識は接話マイクより困難
- マルチモーダル入出力機能
 - 画像, センサ入力との統合
 - 物理実体の動作(ジェスチャーなど)とロボット発話との統合
- 対話ロボットのソフトウェアアーキテクチャ
 - ロボット知能の中の音声対話機能の実現
 - 音声対話研究との融合

Copyright © 2005 Honda Research Institute Japan Co., Ltd.

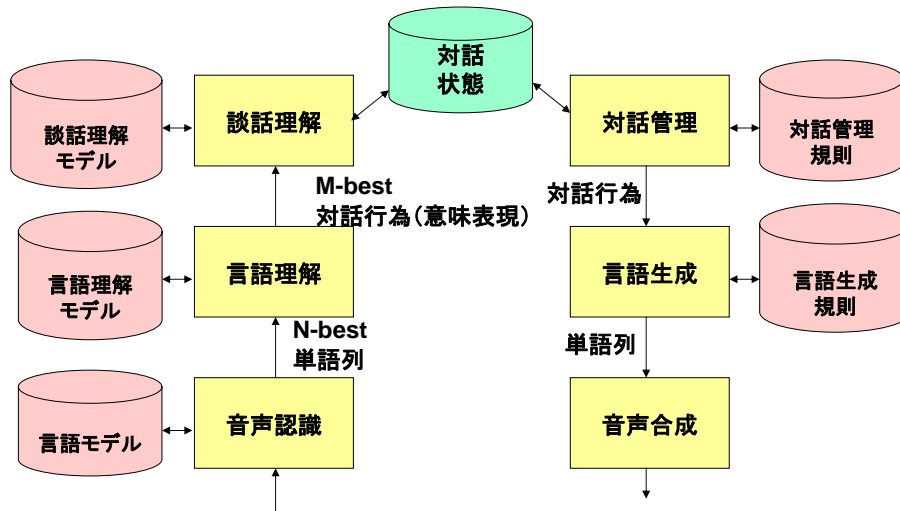
10

音声対話インタフェース研究

音声対話インタフェースの応用

- 電話応答システム
 - e.g. ボイスポータル
<http://www.ntt.com/v-portal/>
 - e.g. 映画の上映時間案内
 - <http://www.voistage.com/ureport/index.htm>
- カーナビゲーションシステム
- 音声案内システム
 - たけまる君(西村, 2004)

音声対話システムの構成



Copyright © 2005 Honda Research Institute Japan Co., Ltd.

13

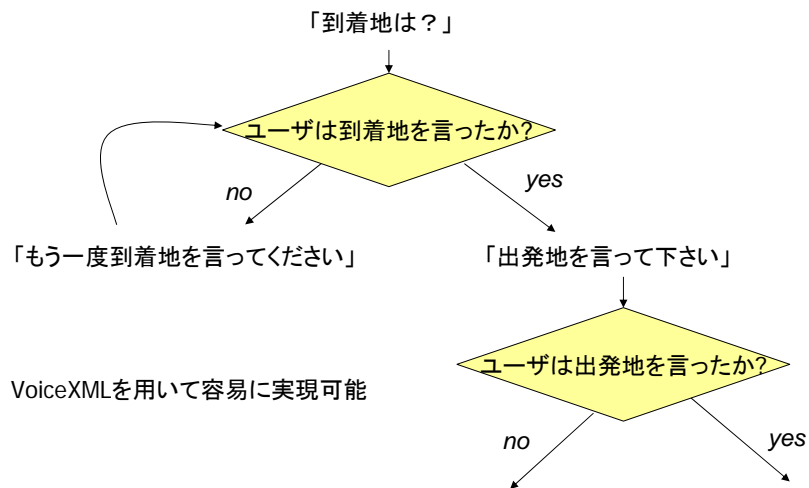
音声対話システムの分類 (cf. 中野・堂坂2002, 駒谷2005)

- 音声認識の語彙サイズ
 - 小語彙 (~100) vs. 中語彙 (~数千)
 - vs. 大語彙 (~数万)
- 言語理解方式
 - キーワード抽出 vs. WFST vs. 構文解析
- 対話状態の表現と対話管理手法
 - 有限状態オートマトン(+ α) vs. フレーム
 - vs. オブジェクト指向 vs. プラン
- タスクの知識
 - 構造を持つDB vs. 大規模テキスト

Copyright © 2005 Honda Research Institute Japan Co., Ltd.

14

状態遷移モデルによる対話管理



Copyright © 2005 Honda Research Institute Japan Co., Ltd.

15

フレームに基づく対話管理

フレームによるユーザ要求意図の表現 + grounding情報

[情報種類: 降水確率 (確認済み)
日にち: 明日 (未確認)
場所: -]

アクション選択:

フレームの状態から次のアクションへの
プロダクション規則を利用

場所が空 ⇒ 場所を聞く
日にちが未確認 ⇒ 日にちを確認

Copyright © 2005 Honda Research Institute Japan Co., Ltd.

16

音声対話システムの発展形

- マルチモーダル対話システム
 - 画像, ペン, ポインティング入力との組み合わせ
(OGI, 豊橋技科大新田研, etc.)
 - 人画像, 画像情報との組み合わせ
 - e.g. Galatea Project
(<http://hil.t.u-tokyo.ac.jp/~galatea/index-jp.html>)
 - Embodied Conversational Agent
(Traum et al; Cassell et al.)
- マルチドメイン対話システム
 - 複数の話題に対処 (O'Neill et al, 2004, etc.)

Copyright © 2005 Honda Research Institute Japan Co., Ltd.

17

音声対話システム研究のホットなトピック

- 音声理解のロバスト化
 - 言語モデルの改良
 - 文脈の利用
- 音声認識誤りの検出と対処
 - 間違っただ確認へのユーザの反応を利用
 - 誤解からの復帰
- 感情認識・生成
- 対話知識の(半)自動獲得
 - 言語モデルや対話戦略などの学習
 - 能動学習, 強化学習など
- 柔軟な話者交代
 - ポーズに頼らない発話終了検出
 - 相槌の認識と生成
- QAシステムとの組み合わせ
 - 大規模テキストコーパスを知識として利用

Copyright © 2005 Honda Research Institute Japan Co., Ltd.

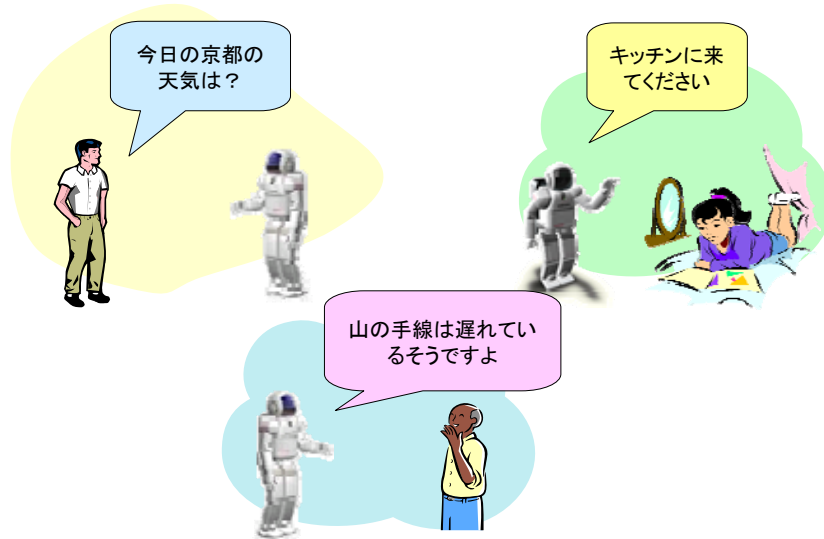
18

ロボットの行動と対話の統合モデル

研究目標

- 対話ロボットの知能ソフトウェアための、
記号レベルモジュールのモデルの構築
 - 人のタスク要求を理解
 - 行動と対話によってタスクを達成
 - 重点を置く機能
 - マルチドメイン対話
 - ドメインの動的変更
 - 割り込みの処理
 - 様々なタスクへ拡張可能

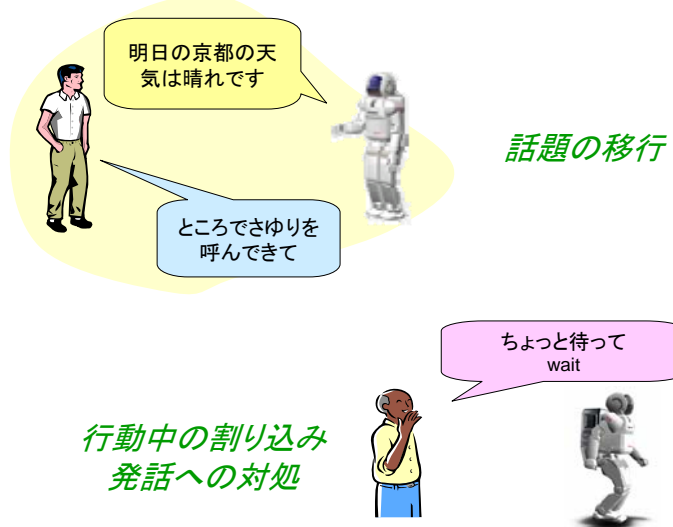
マルチドメイン対話



Copyright © 2005 Honda Research Institute Japan Co., Ltd.

21

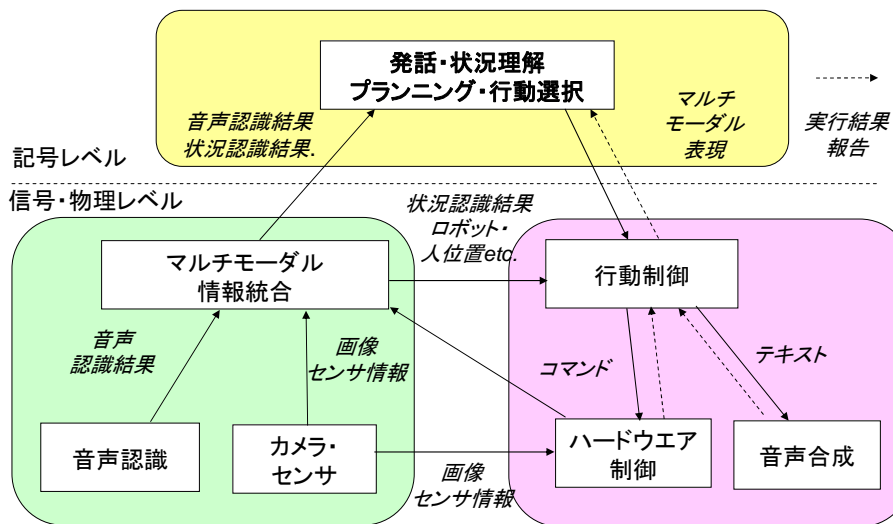
正確でスムーズなコミュニケーション



Copyright © 2005 Honda Research Institute Japan Co., Ltd.

22

対話ロボットのアーキテクチャ



Copyright © 2005 Honda Research Institute Japan Co., Ltd.

23

コミュニケーションロボットのアーキテクチャ

- コミュニケーションロボットでは、発話と行動は特に区別なく設計されている
 - Situated Module
 - Robovie (石黒 2005)
 - Tree-structured behavior modules
 - QRIO (Hoshino 2004)
 - 「やり取り」によって情報を正確に授受する (grounding) プロセスを行うわけではない (その必要がない)

Copyright © 2005 Honda Research Institute Japan Co., Ltd.

24

先行研究: Jijo-2ロボット (産総研 Asoh et al. 2001)

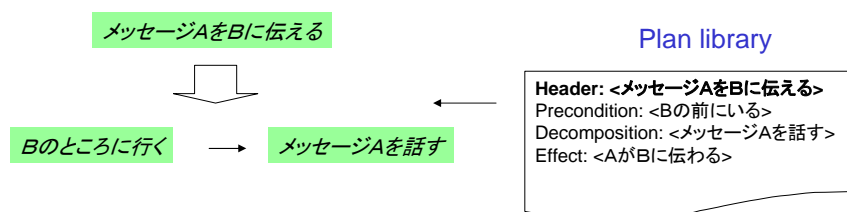
- 行動と対話を統合したタスク遂行が可能
 - 対話管理による正確な理解
 - 複数のドメインの対話
 - 行動中に割り込み発話に反応
- 課題
 - 対話戦略が固定
 - 決まったタイプの対話しかできない
 - 対話と行動の組み合わせに制限
 - タスクの種類が限定



写真提供: 産業技術総合研究所

課題

- ロボットプランニングと対話管理技術の統合
 - 対話と行動の両方を用いてタスクを達成
- タスクを遂行するロボットは単純な状況依存の行動選択だけではなく、階層的なタスクプランニングが必要
- 様々なプランニングストラテジをどう統合するか？



提案法 (Nakano et al. IROS2005)

MEBDP (Multi-Expert-Based Behavior and Dialogue Planner)

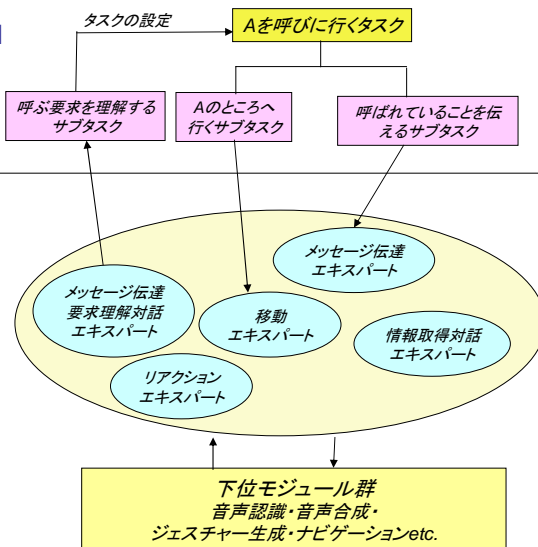
タスクプランニングのレベル

アクション選択のレベル

特定の種類のサブタスクを遂行するのに特化したエキスパートを適宜アクティベート

いろいろなタイプの対話が可能
要求理解からリアクションまで

人間の発話に応じて
エキスパートを動的に変更
(割り込みや話題変更に対処)



Copyright © 2005 Honda Research Institute Japan Co., Ltd.

27

エキスパートの種類

要求理解エキスパート

- 人間の要求を理解する
 - フレームベース対話管理が有効
- 例: 天気予報ドメインにおける
要求理解エキスパート

情報取得対話エキスパート

- プランニングに必要な情報を対話によって得る
- 例: 人位置情報取得
対話エキスパート

情報提供エキスパート

- 人間に情報を提供
 - 割り込み・問い返しに対処できるマルチモーダル発話プランニング
- 例: 天気予報ドメインにおける
情報提供エキスパート

物理動作エキスパート

- 音声インタラクションを伴わない動作の系列をプランニング
- 例: 移動エキスパート

リアクションエキスパート

- 人の発話に1アクションで反応するエキスパート
- 例: 挨拶エキスパート

Copyright © 2005 Honda Research Institute Japan Co., Ltd.

28

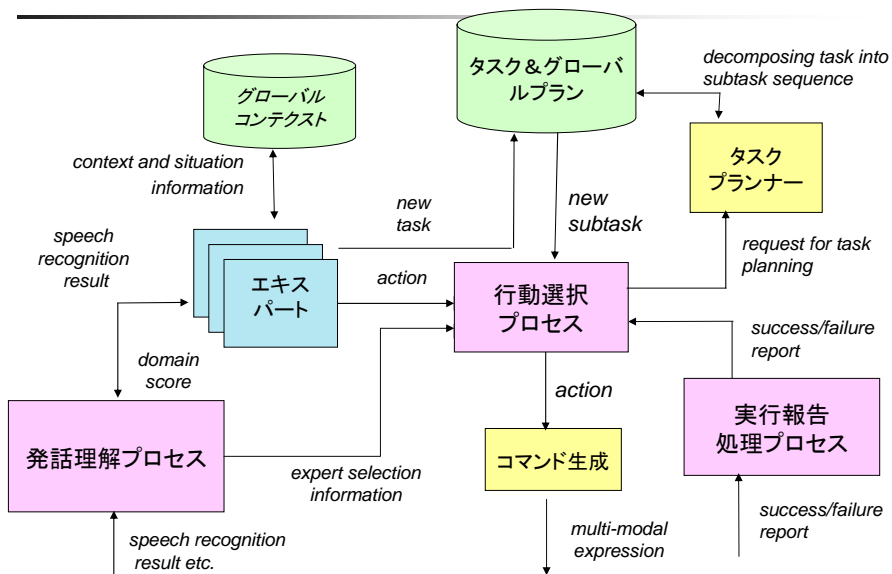
各エキスパートの処理

- エキスパートは内部状態を持つ
 - 人間の要求, 行動の結果 etc.
- エキスパートのインタフェース(オブジェクトのメソッド)
 - understand
 - 発話を理解して内部状態を変更
 - 自分のドメインの発話である確率を推定
 - 音声対話システムの言語・談話理解に相当
 - select-action
 - 内部状態に基づいて発話・行動を選択
 - detect-interruption
 - 直前の発話がロボット発話・行動への割り込みになっているか判断
 - handle-interruption
 - 割り込みへの対処法を決定

Copyright © 2005 Honda Research Institute Japan Co., Ltd.

29

モデルの構成



Copyright © 2005 Honda Research Institute Japan Co., Ltd.

30

実装

- MEBDPを対話ロボットのツールキットとして実装
 - エキスパートを実装するためのテンプレート(abstract class)を用意
 - フレームベース要求理解エキスパートテンプレート
 - 物理行動エキスパート用テンプレート
 - etc.
 - 言語理解部
 - 有限状態トランスジューサ
 - 言語生成部
 - テンプレートベース生成

デモシステム

- 入力: 音声認識システムJulian (Lee et. al, 2002)の出力
 - ネットワーク文法, 約400語, 接話マイク
- 出力: テキスト&物理行動コマンド
 - NTT-IT製 FineVoice 音声合成
 - Honda ASIMOと結合された超音波タグベースのナビゲーションとジェスチャー制御モジュール
- タスク
 - 天気予報, 列車の運行情報などの提供
 - 人を呼びに行く
 - etc.

提案モデルに関する今後の課題

- 発話に基づくドメイン選択の精度向上
- タスクプランニングの改良

ロボット対話の今後の課題

音声認識の向上

- 実環境での音声認識
 - ロボット聴覚 (音源分離, 音源定位, Nakadai et al. 2002)
 - 2005年10月の人工知能学会AIチャレンジ研究会
 - 状況認識の利用
 - e.g. 音声認識がうまくいかないような雑音状況を検出してユーザに伝える (NEC Papero)
 - パラ言語情報の利用
 - 韻律を使って肯定・否定を認識
 - 早稲田大学 小林研 Robisuke (小林, 2005)

Copyright © 2005 Honda Research Institute Japan Co., Ltd.

36

マルチモーダル出力

- 情報伝達における自然性／効率性
 - ジェスチャー, 韻律, その他を利用
 - 例: 頭部色と韻律を用いた感情表現 (Ariyoshi et al. 2004)
 - 頭部色があると感情認識率が上がる
- 設計を容易に
 - MPML-HR (Kushida et al. AAMAS2005 WS)

Copyright © 2005 Honda Research Institute Japan Co., Ltd.

39

MPML-HR スクリプト

```
<mpml>
<head>
  <title> SAMPL SCRIPT </title>
  <agent id="HR" character="ASIMO" />
</head>
<body>
  <page ref="1.html">
    <move x="0" y="500" />
    <play act="Greet" />
    <speak> それでは始めます。</speak>
    <move x="450" y="300" />
    <play act="GestureRight" />
    <speak> MPML は マルチモーダルプレゼンテーションマークアップランゲージの略です。</speak>
    <move x="0" y="0" />
  </page>
</body>
</mpml>
```

- Play <play act=" " />
 - 特定の動作をさせる
 - 現在37種類、増やすことも可能
- Move <move x=" " y=" " z=" " />
 - 指定した座標へ移動
- Speak <speak> </speak>
 - テキストの内容を話す
- Emotion <emotion> </emotion>
 - 音声合成を適切に変化させる
 - 22種類の感情
- Point <point x=" " y=" " />
 - スクリーンを指し示す
- その他
 - <mpml></mpml>, <head></head>, <body></body>, <page ref=" ">, <title></title>, <agent id=" " character=" " />

ute Japan Co., Ltd.

41

物理世界での言語理解・生成

- 物理世界での言語理解・生成 (東工大 田中・徳永研)
 - 言語表現の曖昧性の解消
 - 物理世界での参照表現の生成
 - 言葉と物理世界とのマッピングの記述
- シンボルグラウンディング
 - 音声・画像を使ったインタラクションから言葉の意味を獲得
 - 岩橋 (2003), 稲邑(2004), etc.

Copyright © 2005 Honda Research Institute Japan Co., Ltd.

44

その他のトピック

- 複数の人間との対話
 - 早稲田 小林研 Robita (松坂他, 2001)
 - 視線を用いて発話を促す
- 未知語の獲得
 - 名前を覚える
 - Sony SDR-4X (青山他, 2003)
- その他必要な機能は多数
 - 音声対話インタフェース研究との融合

おわりに: 自然言語処理への期待

- 対話ロボットにおいて自然言語処理+記号処理AIは重要な役割
 - 記号処理モジュールが人とのインタラクションに基づき, 全体の動きを決定
 - 音声認識が不十分な部分をカバー
- 自然言語処理+記号処理AI研究者がロボット研究をするのが容易になっている
 - ハードウェアの進歩
 - ソフトウェアのモジュール化
RTミドルウェア, 商用/フリーの音声認識・合成

参考文献(順不同)

- 瀬名秀明 (2001). ロボット21世紀. 文春新書.
- 長田正 (2005). ロボットは人間になれるか? PHP.
- 梅谷陽二 (2005). ロボットの研究者は現代のからくり師か? オーム社
- Nakano et al. (2005). A Two-Layer Model for Behavior and Dialogue Planning in Conversational Service Robots, Proc. IROS2005.
- Asoh et al. (2001). Jijo-2: An office robot that communicates and learns, IEEE Intelligent Systems vol. 16, no. 5, pp. 46--55, 2001.
- 日経トレンドィ (2005). 先端ロボット技術はここまで来たー「一家に一台」が現実に! 日経トレンドィ2005年11月号
- 青山 他 (2003). ユーザ固有の情報を獲得・再利用するロボットでの音声対話. 人工知能学会SIG-SLUD-A301.
- 西村 他 (2004). “実環境研究プラットフォームとしての音声情報案内システムの運用”, 電子情報通信学会論文誌, Vol.J87-D-II, No.3, pp.789-798.
- 中野・堂坂 (2002). 音声対話システムの言語・対話処理. 人工知能学会誌 17(3) pp. 271—278.
- 駒谷 (2005). 音声対話システムにおける対話のモデル. 人工知能学会 SIG-SLUD-A501.

- Hoshino et al. (2004). Behavior description and control using behavior module for personal robot. Proc. ICRA-2004.
- Lee et al. (2002). Julius - an open source real-time large vocabulary recognition engine. Proc. Eurospeech.
- 石黒他(2005). コミュニケーションロボット. オーム社
- Nakadai et al. (2002). Exploiting Auditory Fovea in Humanoid-Human Interaction. Proc. AAAI.
- Tokunaga et al. (2003). Bridging the Gap between Language and Action. Proc. IVA 2003.
- 岩橋 (2003). 言語コミュニケーションのための機械学習. 人工知能学会誌 18(5).
- 稲邑他 (2004). ミメシス理論に基づく見まね学習とシンボル創発の統合モデル. 日本ロボット学会誌, Vol.22, No.2, pp.256--263.
- 小林(2005). パラ言語の理解・生成機能をもつ会話ロボット. 信学技法 NLC2005-31/PRMU2005-58.

-
- 松坂他 (2001). "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DII, Vol.J84-D-II, No.6, pp.898-908.
 - Ariyoshi et al. (2004): "Multimodal expression for humanoid robots by integration of human speech mimicking and facial color", Proc. Interspeech.
 - O'Neill et al. (2004). Cross Domain Dialogue Modelling: An Object-Based Approach. Proc Interspeech.
 - M. McTear (2004). Spoken Dialogue Technology. Springer.