

# LCore: 言語と動作によるコミュニケーションを学習するロボット知能化技術

岩橋 直人  
情報通信研究機構 (NICT)



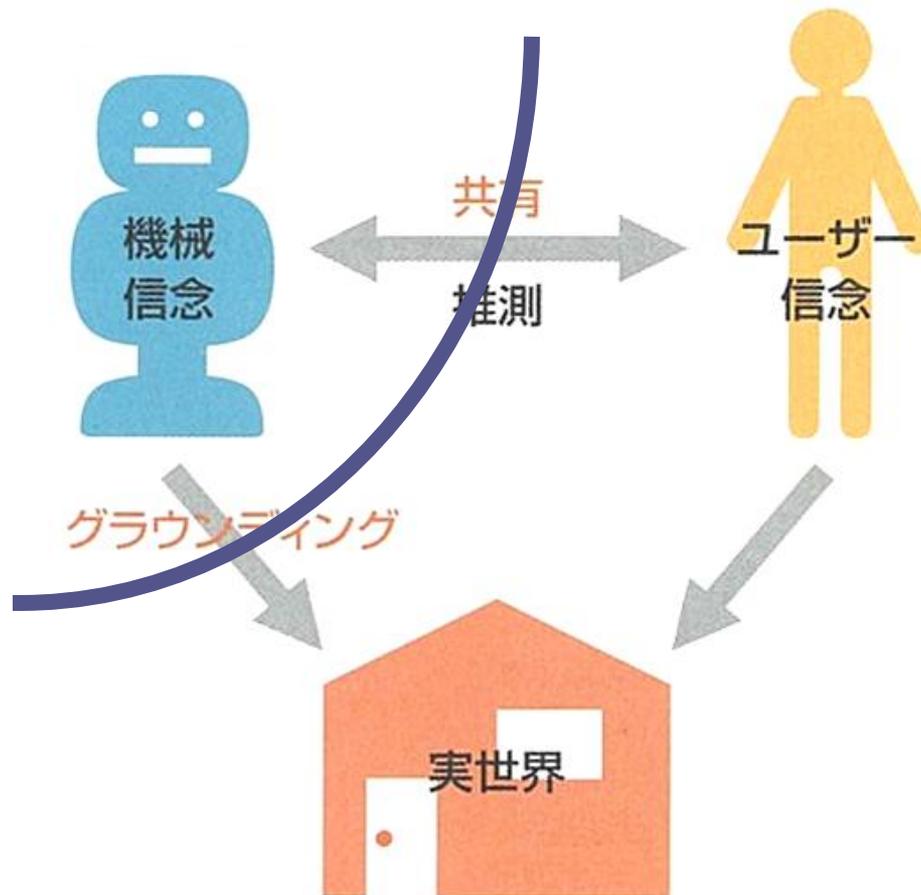
# Agenda

1. はじめに
2. コミュニケーション学習手法LCoreの概要
3. 音声言語学習手法の例
  1. モデル選択に基づく文発話からの単語学習
  2. 発話理解の学習
4. まとめ

# 背景

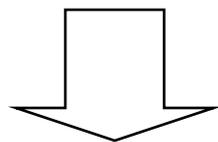
- 日常生活支援ロボットの対話機能は現状ではまったく不十分である
- ロボットの対話技術はとても難しい!
- どうして?
  - 従来の言語処理は閉じた記号系に基づく
  - 実世界に関する共有信念を形成できない
  - 「いつものあれ持ってきて」
  - 「これを引き出しにもどしておいて」

# ロボットに必要な3つの要件



# アプローチ

- 知覚運動情報による学習は必須
- 幼児による言語獲得
- コミュニケーション能力を学習するロボット



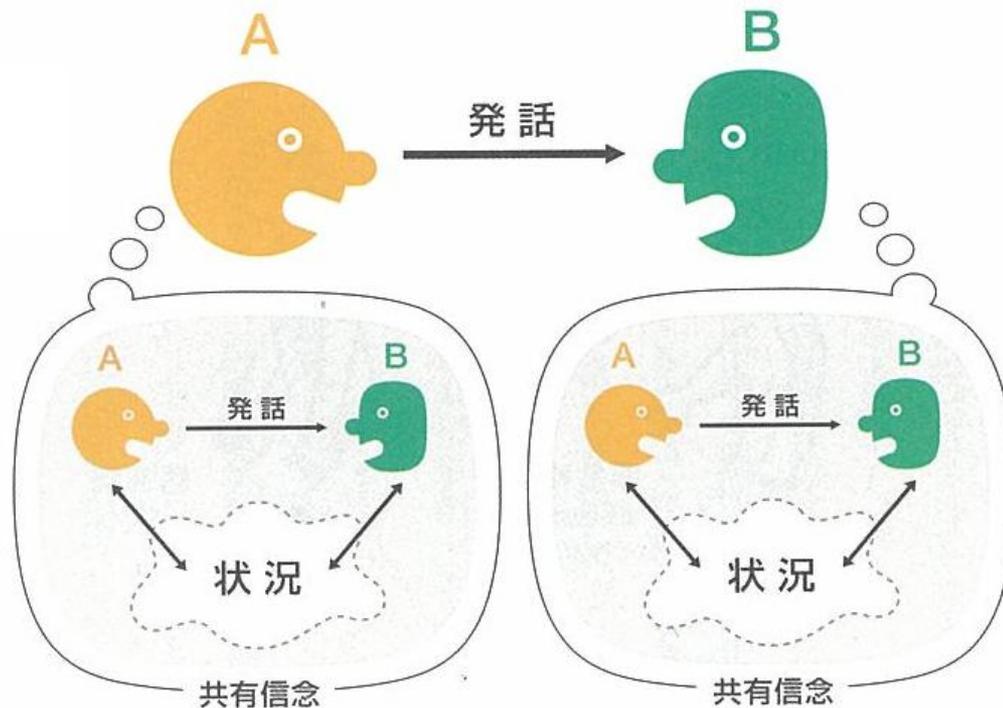
コミュニケーション技術への発達的アプローチ

# 共有信念に基づくコミュニケーション

## 対話例

翔太: コーヒーを飲むかい

美咲: コーヒーを飲むと目が覚めるわ



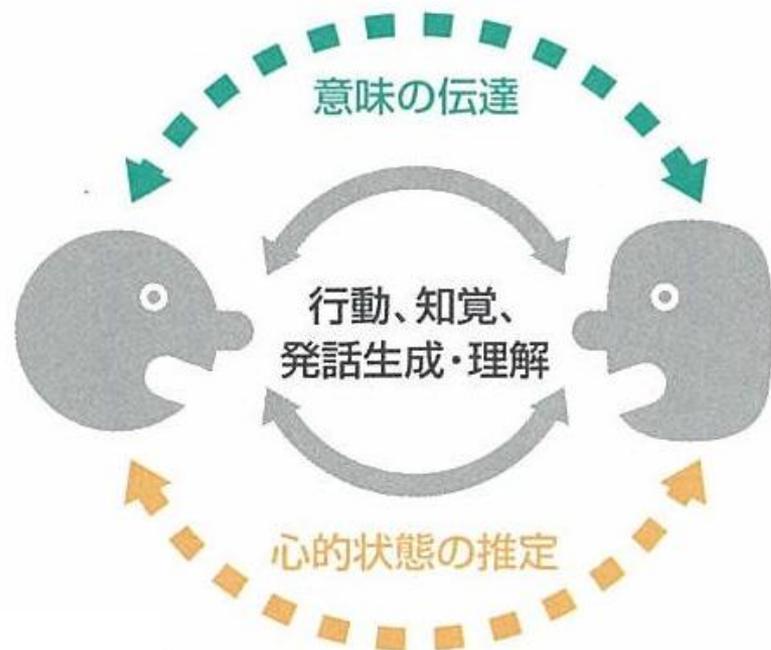
# 心的状態の推定

## 対話例

話者A： おはよう。昨夜、見た？

話者B： うん、見たよ。なかなか良かったな。

話者A： そうだね。



# Agenda

1. はじめに

2. コミュニケーション学習手法LCoreの概要

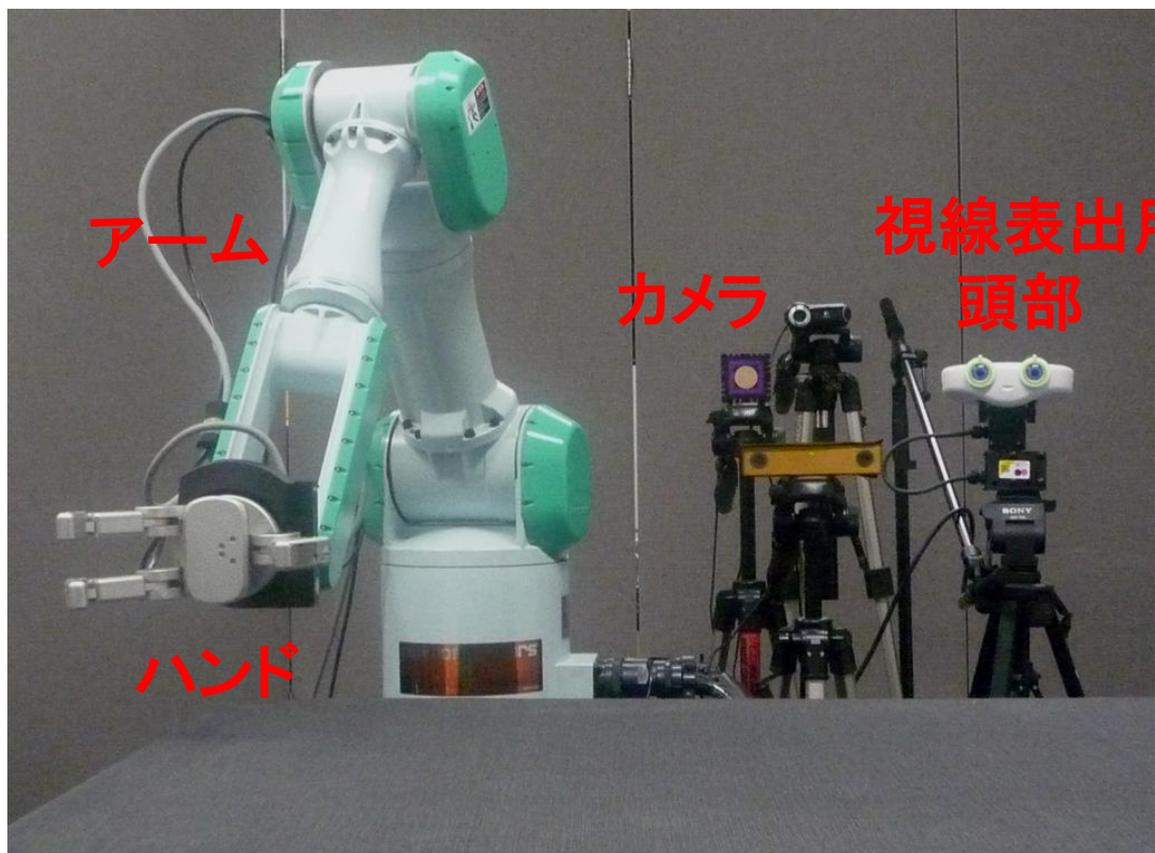
3. 音声言語学習手法の例

1. モデル選択に基づく文発話からの単語学習

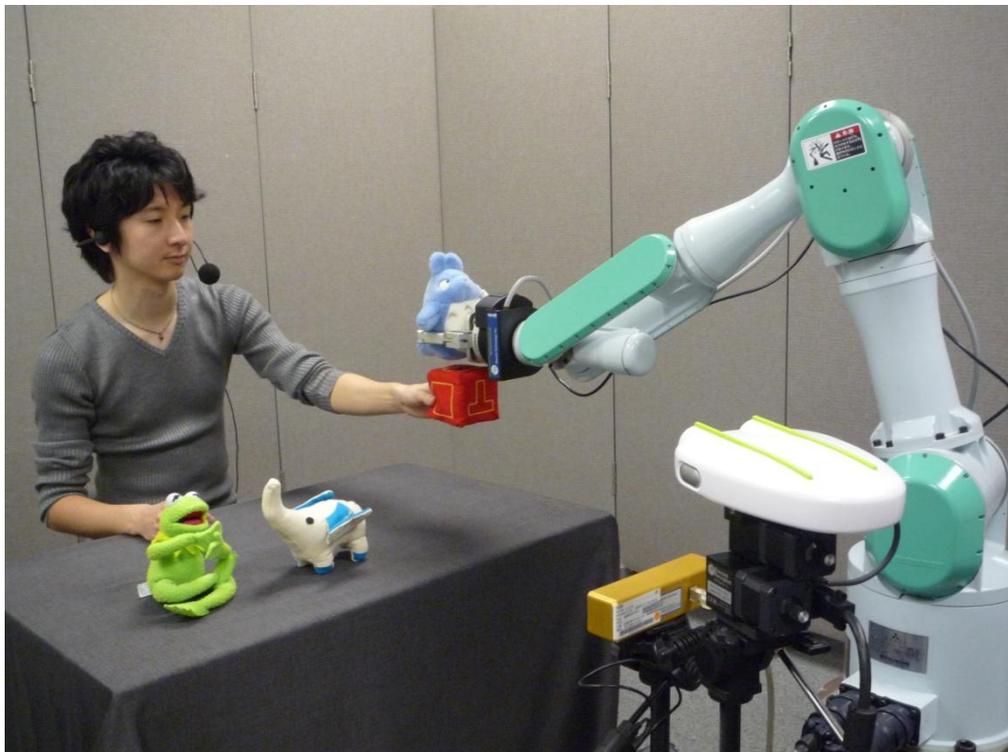
2. 発話理解の学習

4. まとめ

# ロボットプラットフォームホーム (1)



## ロボットプラットホーム (2)



# 現在LCoreができること

音韻学習

単語学習

文法学習

語用法学習

発話生成・理解

ロボットへの発話の検出

質問応答の学習

物体画像概念学習

物体抽象概念学習

物体操作動作学習

協調動作の役割交代模倣

ユーザ信念の推定

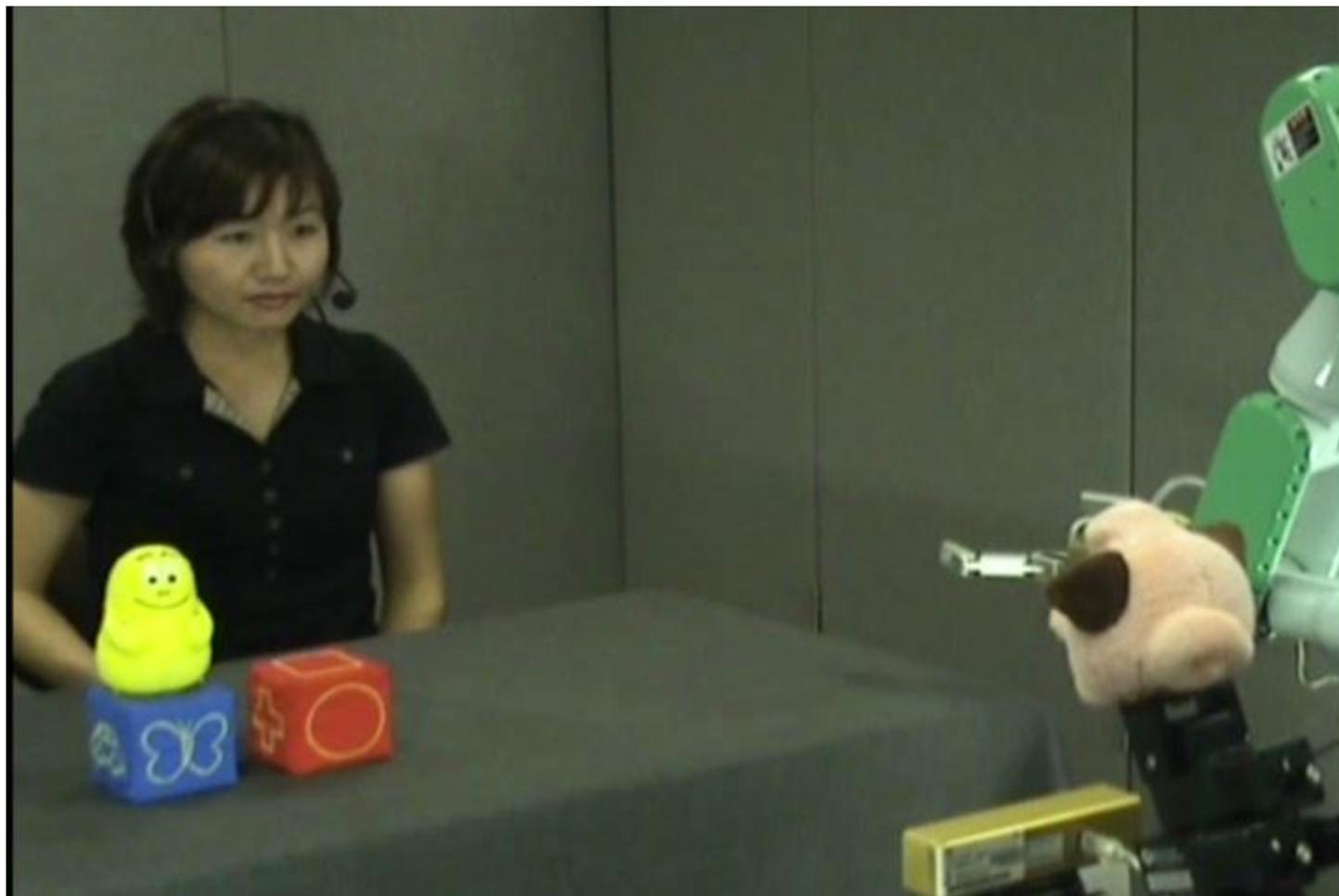
# エピソード例 一発話理解誤り



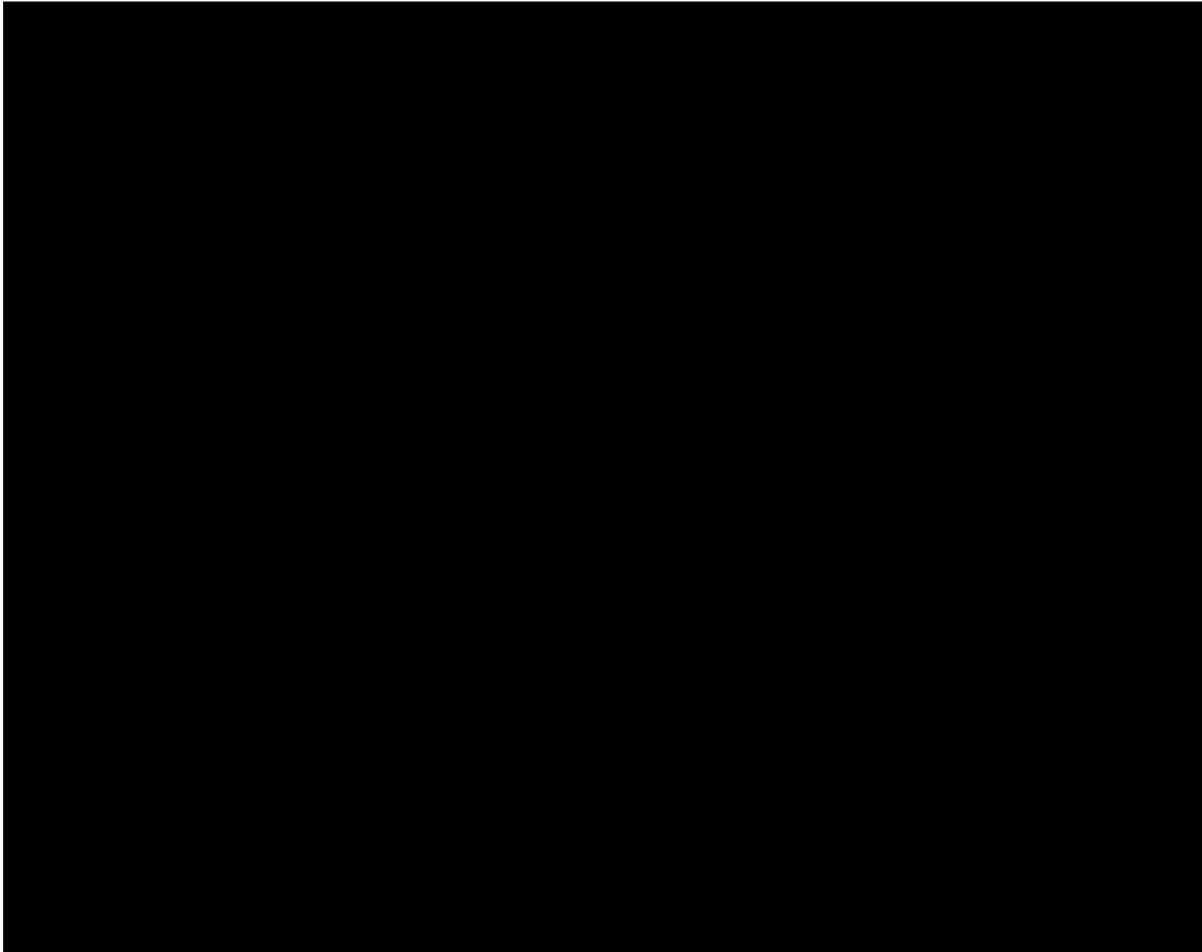
# エピソード例 一発話理解誤り



# エピソード例 一発話理解正解一



# エピソード例　—発話生成—



# Agenda

1. はじめに
2. コミュニケーション学習手法LCoreの概要
3. 音声言語学習手法の例
  1. モデル選択に基づく文発話からの単語学習
  2. 発話理解の学習
4. まとめ

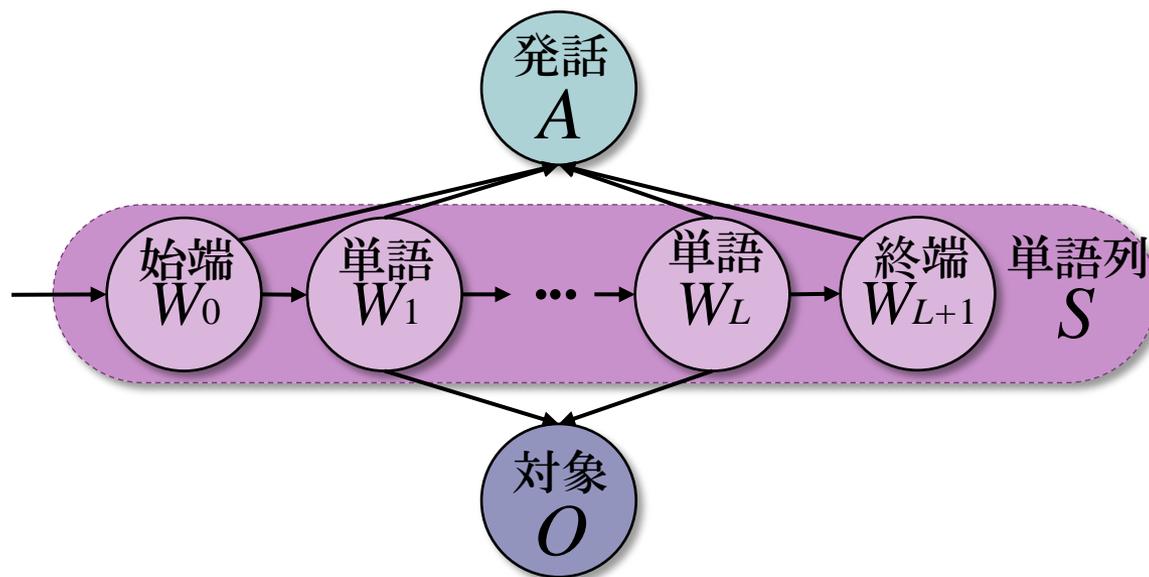
# 問題設定

単語や文法の知識を持たないロボットが  
**文音声と対象のペア** から単語とその意味を学習する



# 発話と対象の共起確率モデル

三種類の確率モデル（音響，文法，語意）を統合



$$\log P(A, O) = \log \sum_S P(A | S) P(S) P(O | S)$$

$$\approx \max_{s \in N_{Best}} \left\{ \omega_1 \log P(A | S; \theta) + \omega_2 \log \prod_{l=0}^{L^s} P(W_{l+1} | W_l) + \omega_3 \log \sum_{l=1}^{L^s} \gamma_l P(O | W_l) \right\}$$

音響
文法
語意

# 提案手法の流れ

音素認識

## 初期単語リストの構築

音声

	音素列
$w_1$	k/o/k/o/w/a
$w_2$	g/a/k/u/s/e
$w_3$	b/e/y/a
...	

単語リスト

部分音素列の種類: 約6000種類  
(60発話)

各部分列の  
前後1モーラの情報量  
が閾値以上なら追加

? かいぎし ?

? かいぎしつ ?

モデル選択

単語リストの項目数: 約200単語

学習データ

# モデル選択による単語リスト最適化

- 1 Bigramの確率が高い2単語を連結
- 2 MDL基準に従って不要な単語を削除する

データのあてはまりの良さとモデルの自由度とのバランスを取る

$$\text{記述長} = -(\text{モデル尤度}) + \frac{\text{自由度}}{2} \log(\text{データ数})$$

$$\text{モデル尤度} = \sum_i^{\text{データ数}} \log P(A_i, O_i)$$

$$\text{自由度} = \{ \text{単語数}^2 + 2 \times \text{単語数} \} + \{ \text{対象数} \times \text{単語数} \}$$

## モデル1の単語

$w_1$  k/o/k/o/w/a

$w_2$  h/o/k/o/w/a

$w_3$  g/a/k/u/s/e

...

VS

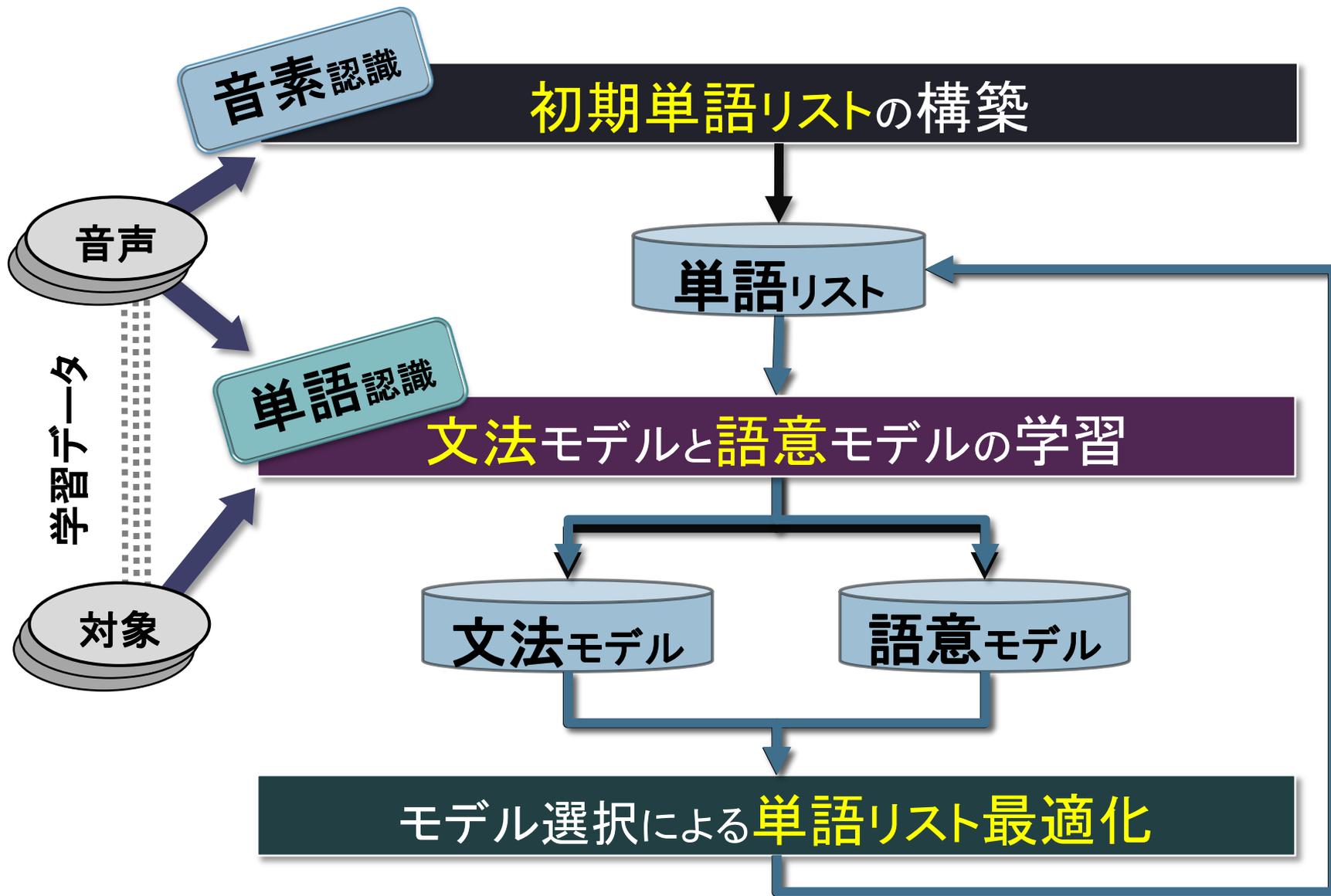
## モデル2の単語

$w_1$  k/o/k/o/w/a

$w_3$  g/a/k/u/s/e

...

# 提案手法の流れ



# 学習用収録音声

キーワード10種類 ✕ 言い回し6種類 = 60発話を収録

対象	キーワード	対象	キーワード
1	会議室の前	6	竹内さんのブースの南
2	辻野さんのブース	7	工作室
3	フロアの真ん中	8	アシモの部屋
4	学生部屋の前	9	スマートルーム
5	お茶飲み場	10	スマートルームの入り口

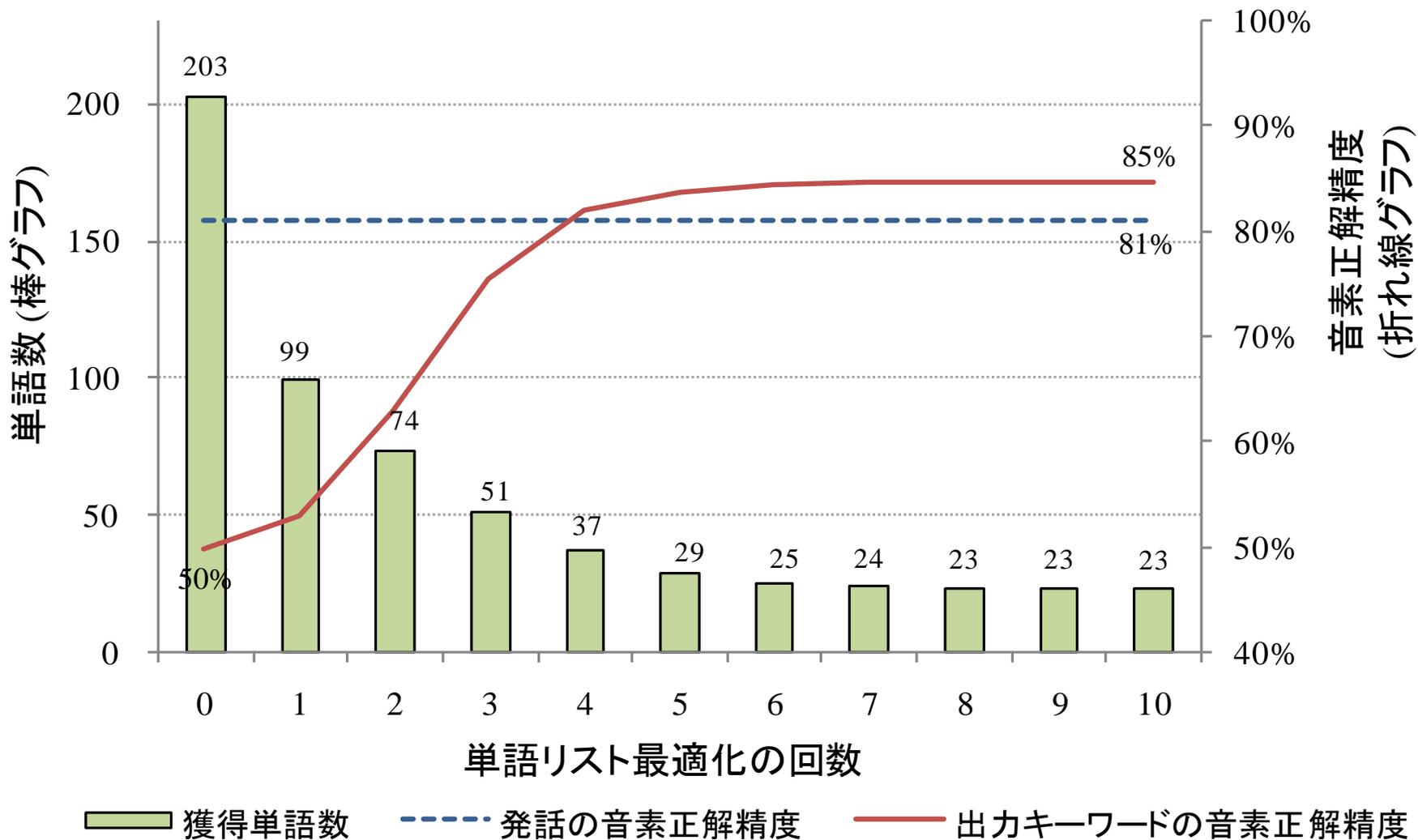


## 言い回しのパターン

この場所は～	ここは～です	この名前は～
～の所に行って	～へお願い	今から～へ行って

音素正解精度は平均 **81 %**

# 実験結果（16名分の平均）



# 実験結果：提案手法で獲得されたキーワード例

正解キーワード	最適化なし	最適化10回
会議室の前	かいすのまえ	かいすのまえ
辻野さんのブース	つじのさ	つじのさうのぶす
フロアの真ん中	なか	ふろあどまんなか 
学生部屋の前	がくせえべや	がくせえべやのまえ
お茶飲み場	おちょ	おちゃのいま
竹内さんのブースの南	み	たきよいつさんのぶすのみなみ
工作室	こおさくしつ	こおさくひつ
アシモの部屋	あしものへや	あしものへや
スマートルーム	む	すもあとるむ
スマートルームの入り口	ち	すまとるむのいいぐち
平均 音素正解精度	<b>43</b> %	<b>85</b> %



単語リストの最適化によって分節誤りが修正される

# 学習中の認識結果の例



発話「この場所はフロアの真ん中」

音素認識結果：こどばしえおあふろあのまんがか

(kodobasheoahuroanomangaka)



初期モデル : / こど / ば / しょわ / ふろあ / ど / まんなか /



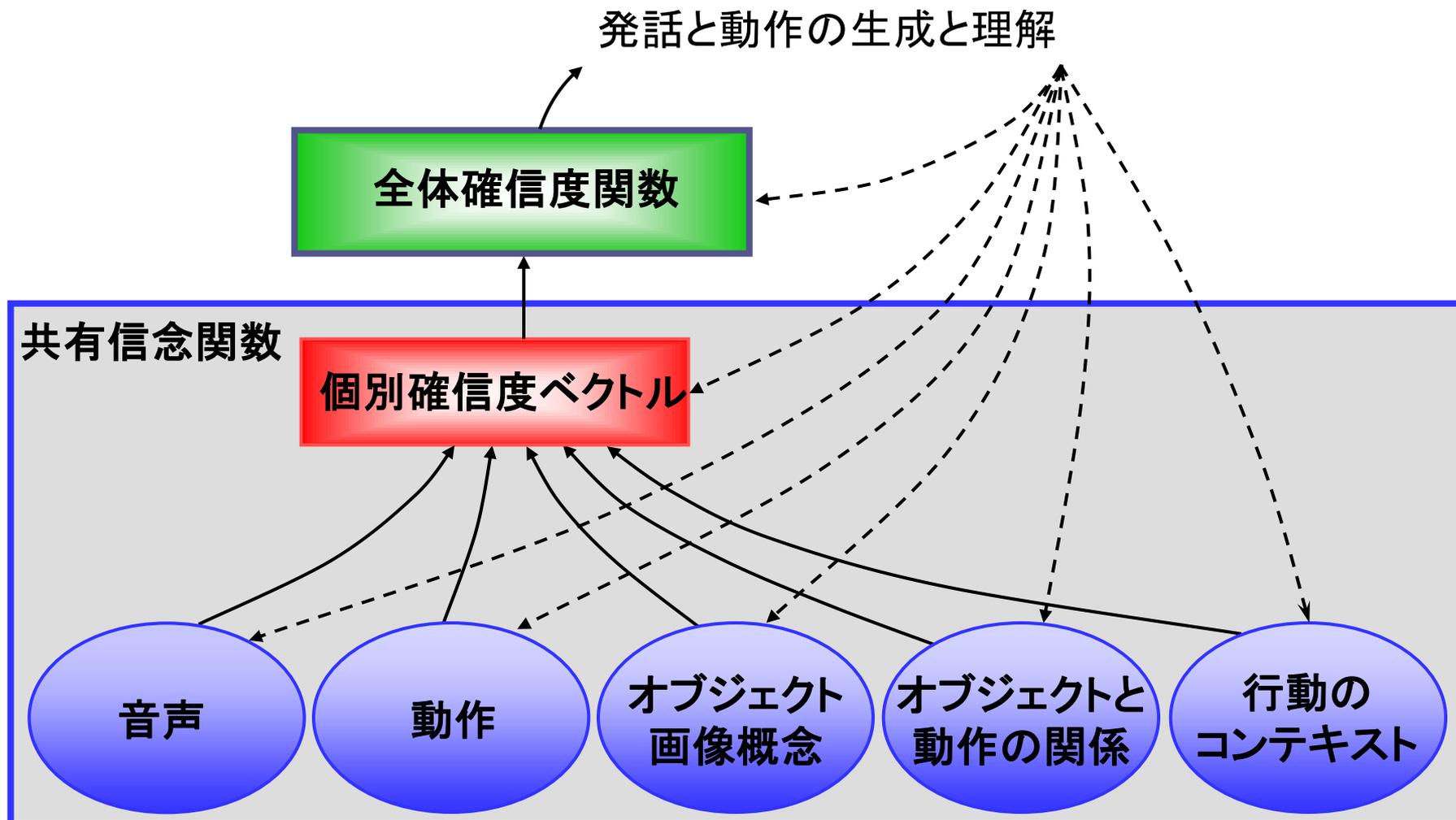
モデル選択1回 : / この / ばしょわ / ふろあどまん / なか /

モデル選択2回 : / このばしょわ / ふろあどまんなか /

# Agenda

1. はじめに
2. コミュニケーション学習手法LCoreの概要
3. 音声言語学習手法の例
  1. モデル選択に基づく文発話からの単語学習
  2. 発話理解の学習
4. まとめ

# ロボットの信念システム



# 共有信念関数

- 信念モジュールと個別確信度から構成
- 出力: 発話  $\mathbf{s}$  と行動  $\mathbf{a}$  関連性の度合い

## 個別確信度

$$\Psi(s, a) = \max_{l, z} \left\{ \begin{array}{l} \gamma_1 \log p(s | z; L, G) \quad \text{音声} \\ + \gamma_2 (\log p(t | W_T; L) + \log p(l | W_L; L)) \quad \text{オブジェクト画像} \\ + \gamma_3 \log p(u | W_M; L) \quad \text{動作} \\ + \gamma_4 \log p(t, l | W_M; R) \quad \text{オブジェクトと動作の関係} \\ + \gamma_5 \log p(t, l | q; H) \quad \text{行動のコンテキスト} \end{array} \right\}$$

# 発話理解

動作



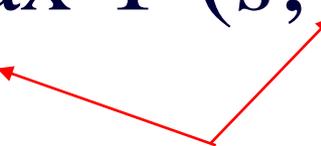
発話



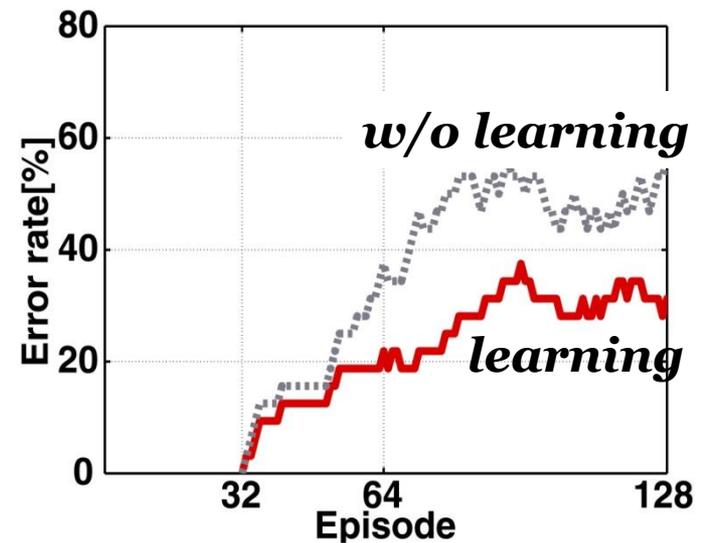
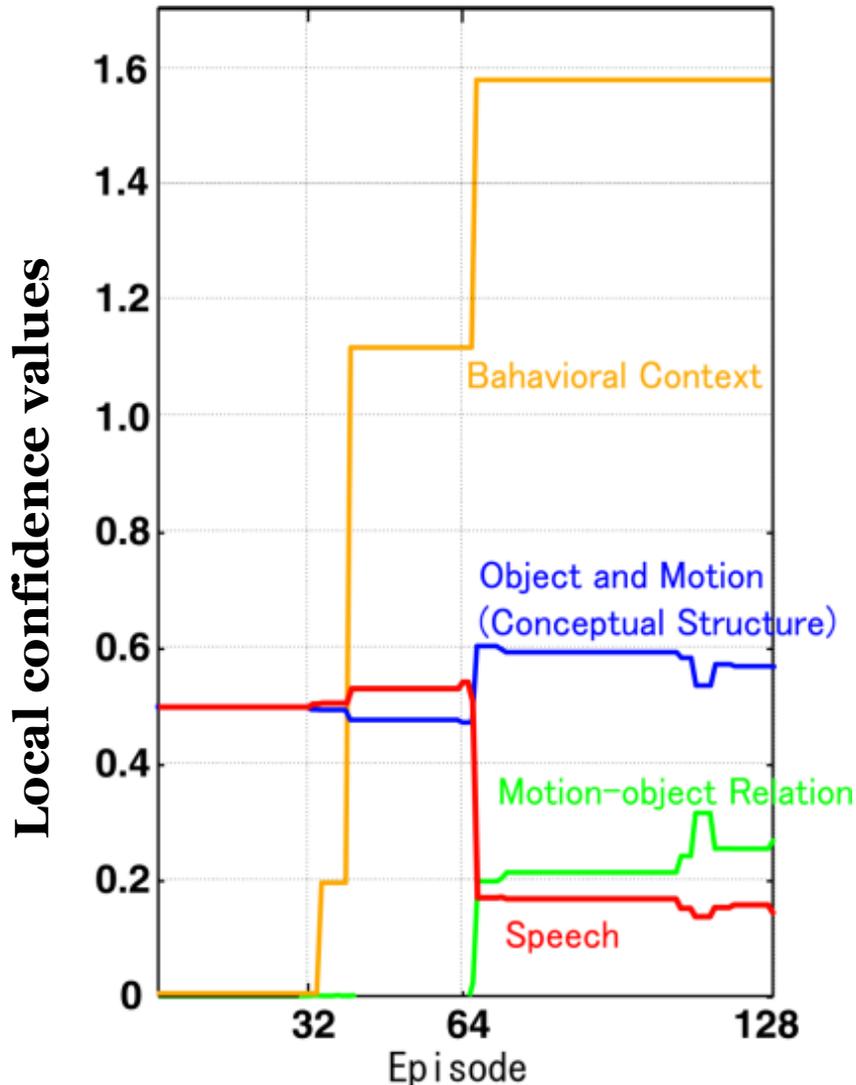
$$a = \arg \max \Psi (s, A)$$

A

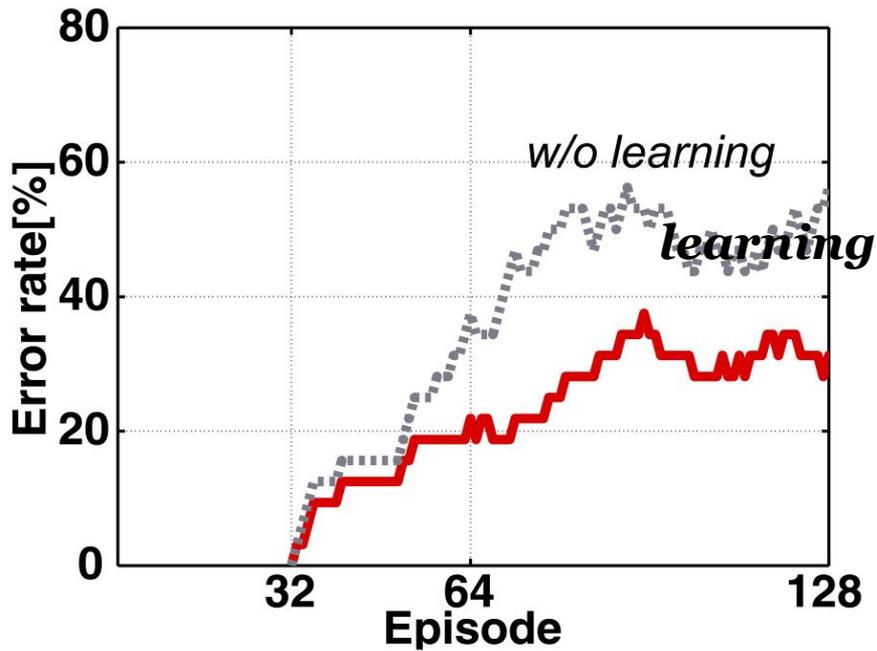
可能な動作



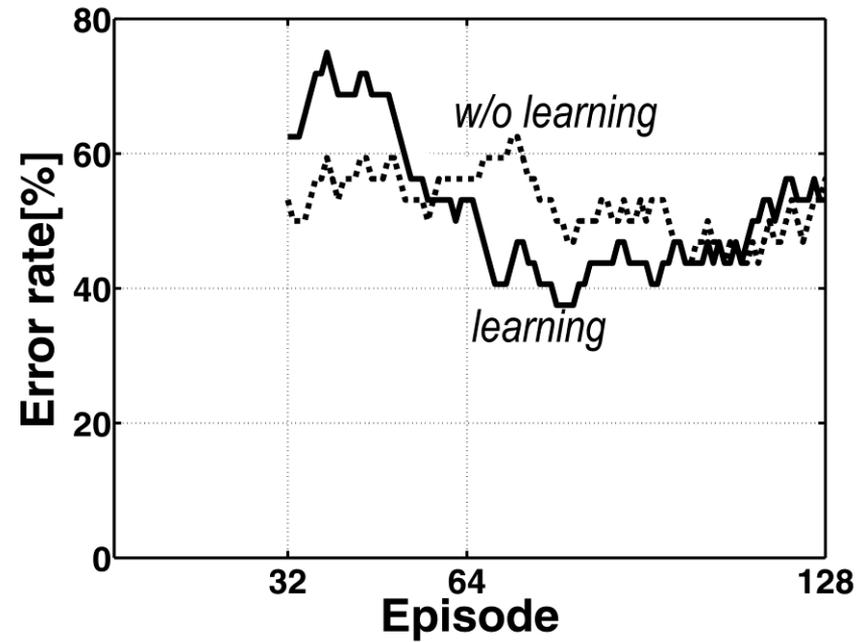
# MCE学習による個別確信度の変化



# カップリングのブレイクダウン



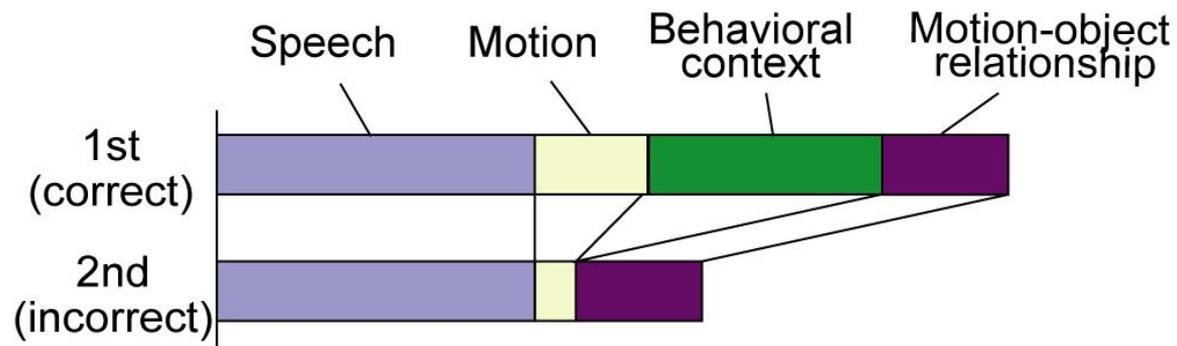
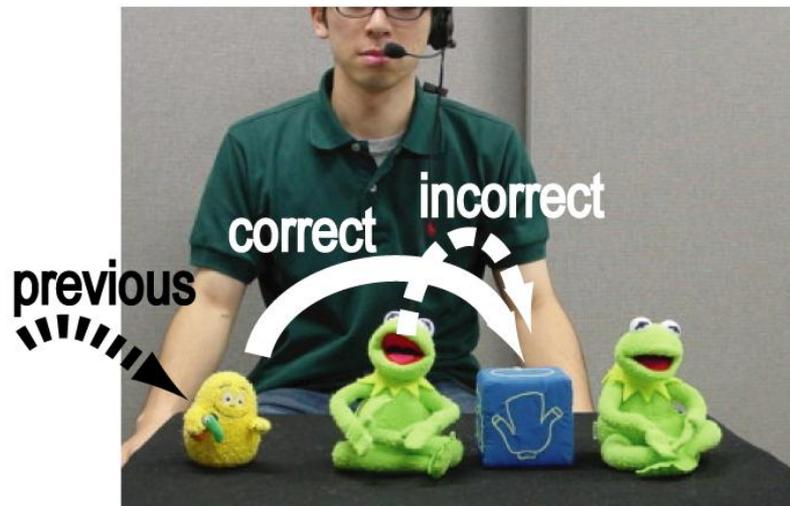
complete → fragmental



fragmental → fragmental

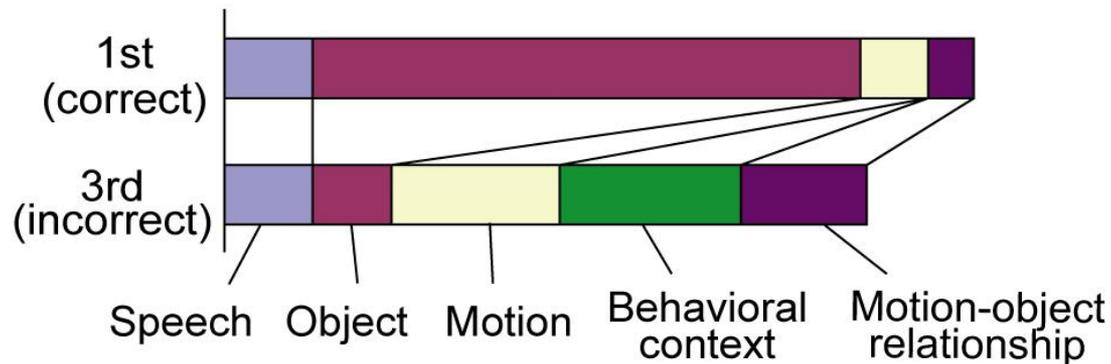
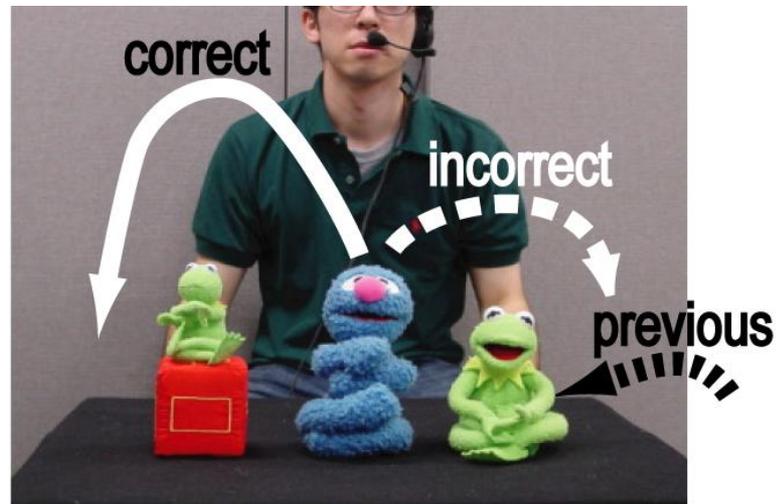
# 発話理解例 (1)

“のせて”



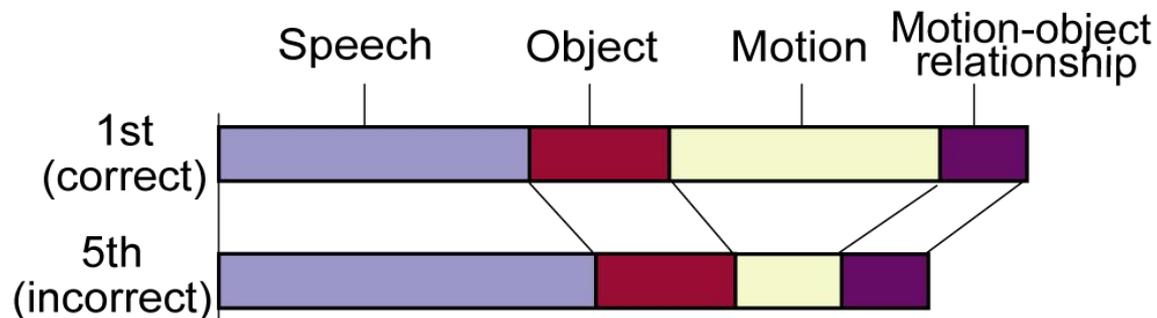
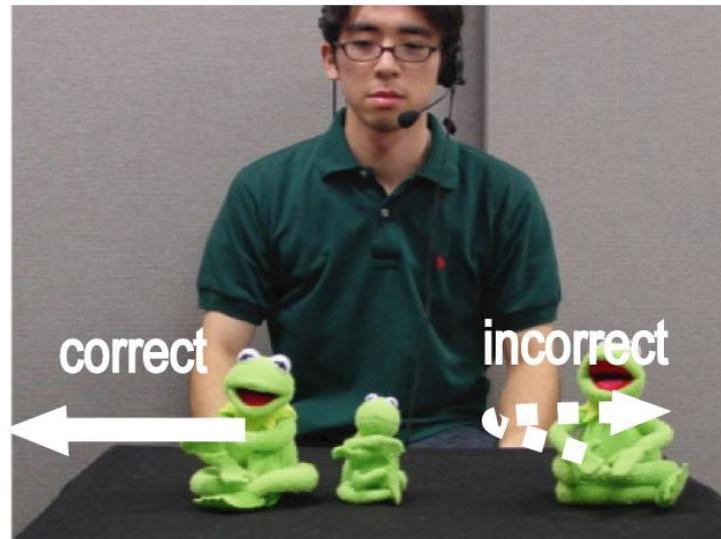
# 発話理解例（2）

“グローバー 小さい カーミット 飛び越えさせて”



# 発話理解例（3）

“大きいカーミット 小さいカーミット 遠ざけて”



# Agenda

1. はじめに
2. コミュニケーション学習手法LCoreの概要
3. 音声言語学習手法の例
  1. モデル選択に基づく文発話からの単語学習
  2. 発話理解の学習
4. まとめ

# まとめ

- 言語獲得ロボットの概要紹介
  - 「いつものあれ持ってきて」を理解する基本技術
    - 実世界におけるユーザとのやり取りを通して、ユーザの意図や状況を適切に理解してコミュニケーションする能力を学習する知能化技術
  - ロボットとの対話技術に必要な3つの要件を充足
    - グラウンディング
    - 拡張性
    - 信念共有
- 音声言語処理技術の例
  - モデル選択に基づく単語学習
  - 発話理解の学習

# これからの展開

- より実用的に
  - 机上からリビングルームへ
  - 移動型双腕ロボットへの実装

ロボカップ・@ホームリーグ  
優勝

電通大・玉川大との共同

- より知的に
  - より制約のない発話からの学習
  - 行為の目的の理解
  - 言語の超越性



ありがとうございました