



ヒューマンコンピュータインタラクション特論 第9回

情報理工学部門 複合情報工学分野

Human-Computer Interaction (HCI) 研究室

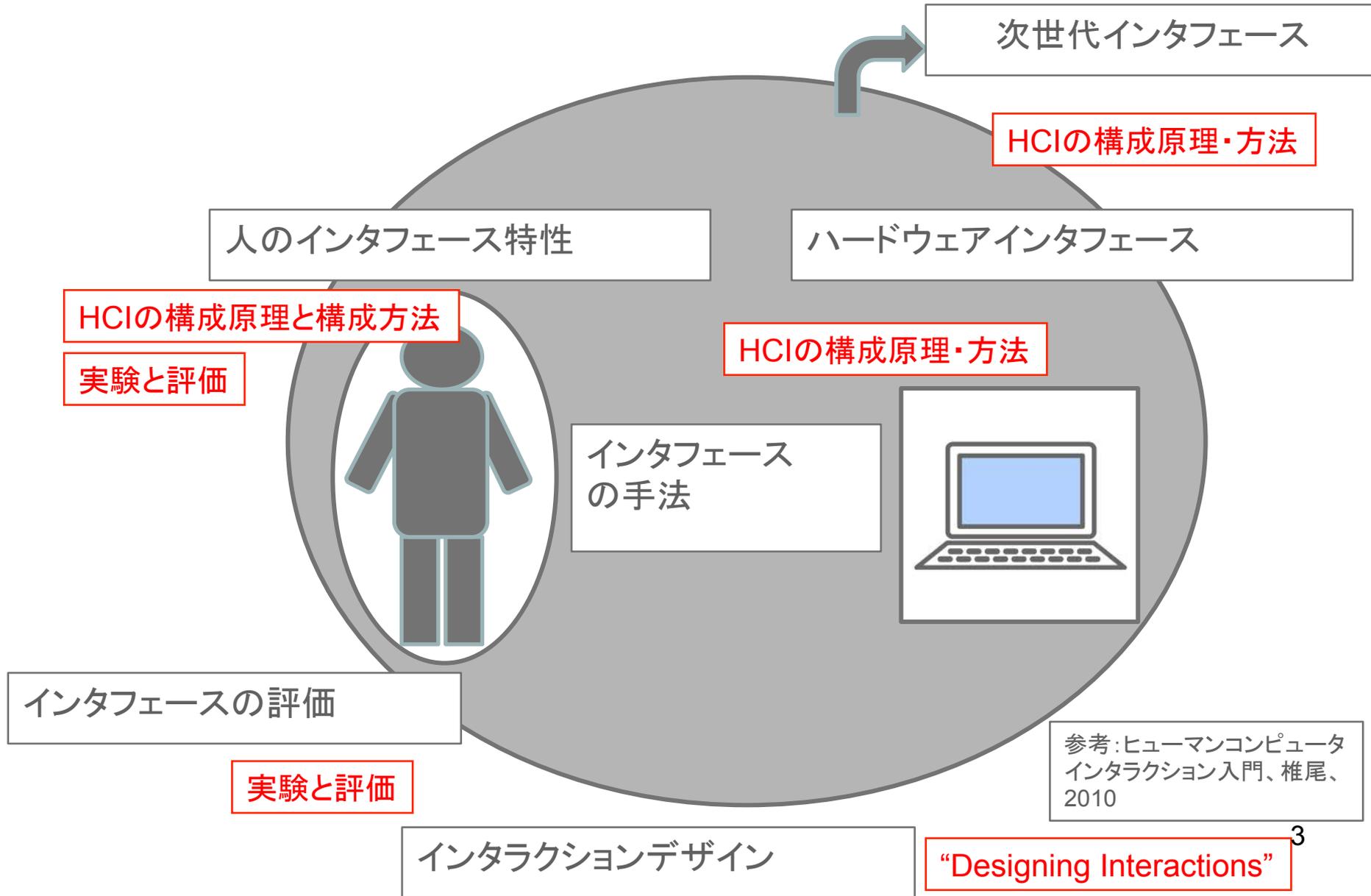
小野哲雄 (8-12室)

tono@ist.hokudai.ac.jp

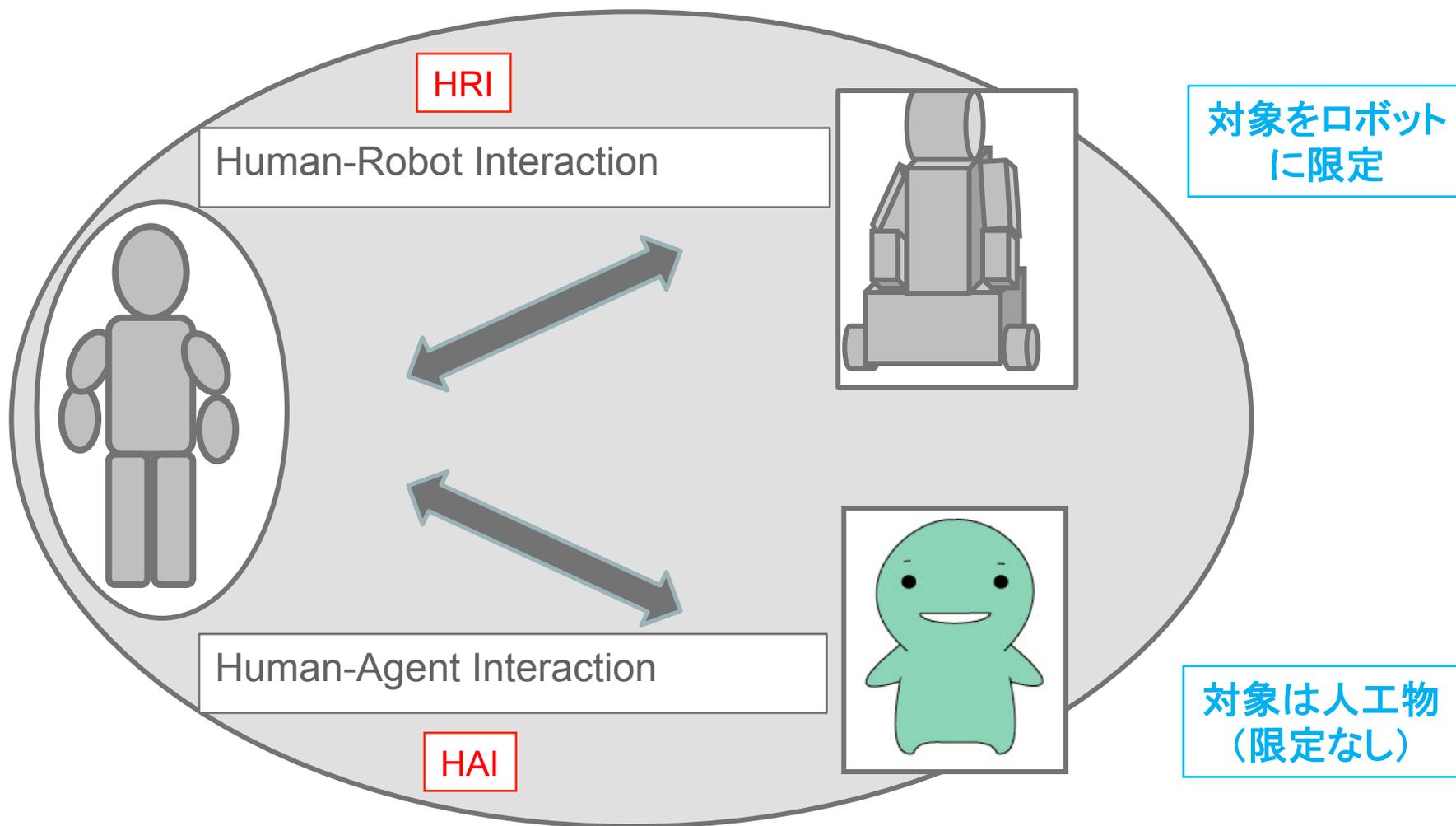
授業計画 (予定)

1. 10/1(火) 4限 (小野) HCI入門
2. 10/4(金) 3限 (小野) HCI概論
10/8(火) 4限 : **[休講]**
3. 10/11(金) 3限 (小野) HCI概論(2) # 論文発表の説明
4. 10/15(火) 4限 (小野) HCIの構成原理と構成方法(1) # 登録開始
5. 10/25(金) 3限 (小野) HCIの構成原理と構成方法(2)
6. 10/31(木) 4限 ← 火曜授業 (小野) **[論文発表1(1)]**
7. 11/1(金) 3限 (小野) **[論文発表1(2)]**
8. 11/5(火) 4限 (小野) ヒューマンロボットインタラクション(HRI) # 説明
9. 11/8(金) 3限 (小野) ヒューマンエージェントインタラクション(HAI) # 登録
10. 11/12(火) 4限 (小野) Predicting Human Decision-Making (1)
11. 11/15(金) 3限 (小野) Predicting Human Decision-Making (2)
12. 11/19(火) 4限 (坂本) HCIにおける実験と評価(1)
13. 11/22(金) 3限 (坂本) HCIにおける実験と評価(2)
14. 11/26(火) 4限 (小野) **[論文発表2(1)]**
15. 11/29(金) 3限 (小野) **[論文発表2(2)]**

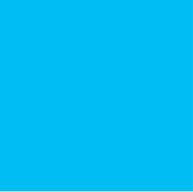
講義の概要（1）：HCI全般



講義の概要 (2) : HRIとHAI



HCIとの相違点：人と「エージェント」のインタラクションデザインを研究
HCIとの共通点：人を含む「開放系」を研究の対象とする



Human-Agent Interaction (HAI) (Human-Robot Interaction (HRI))

エージェントを用いた対話システム

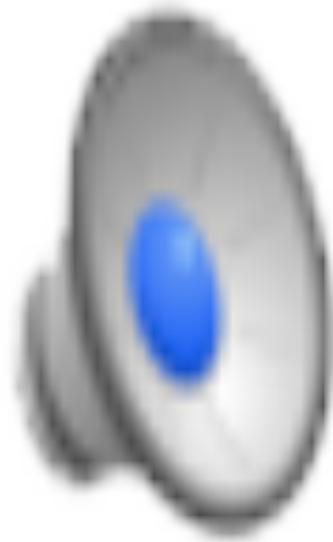
- 研究事例
 - PUT-THAT-THERE (Bolt 1980)
 - Rea (Cassell 2000)
 - Situated Interaction (MSR 2009, 2014)
 - Mentalizing:
 - Rational quantitative attribution of beliefs, desires and percepts in human mentalizing
(*Nature human behavior* 2017)
- 技術的な要素
 - POMDP, Generative model, (Bayesian inference)

Human-Agent Interactionとは？ (1)

- 会話エージェントの研究を概観
 - 研究の流れ
 - 自然言語対話システム ➡ 音声対話システム ➡ マルチモーダル対話システム ➡ 擬人化エージェント
 - 研究の歴史
 - 1967: ELIZA [Weizenbaum] 表層的な対話を実現
 - 1972: SHRDLU [Winograd] トイワールド内の言語理解
 - 1973: LUNAR [Woods] 自然言語解析. 機械が人間へ近づく
 - 1980: HEARSAY-II [Erman] 音声対話に基づくタスク処理
 - 1980: PUT-THAT-THERE [Bolt] 音声(代名詞)で操作を指示
 - 1987: The Knowledge Navigator [Apple] コンセプトムービー
 - 1994: CASA [Nass] “Computers Are Social Actors”
 - 1994: Believable agents [Bates] 生き物らしいエージェント
 - 1996: The Media Equation [Reeves & Nass] メディアの等式
 - 1997: Persona [Ball] ToolからAssistive interfaceへ
 - 2000: Rea [Cassell] Embodied Conversational Agents

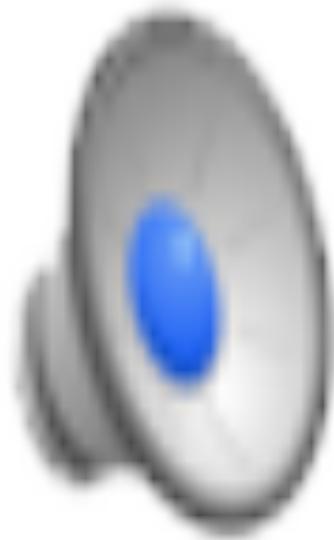
Human-Agent Interactionとは？ (2)

- PUT-THAT-THERE [Bolt 1980]



Human-Agent Interactionとは？ (3)

- Rea [Cassell 2000] (ビデオ 1'00 - 2'15)



Human-Agent Interactionとは？ (4)

- 『HAI研究のオリジナリティ』 (山田 09 参照)
 - ・ 人-エージェント、人-ロボット、人-エージェント-人のインタラクション
 - ・ HAI研究は開放系の研究が特徴
 - ・ HAIではインタラクションの対象をコンピュータやロボットに限定しない
 - ・ 「認知的インタラクションデザイン学」の重要性
 - ・ 事例：
 - アプライアンスのエージェント化
 - ロボット(AIBO)が人とのインタラクションを学習
 - エージェントの内部状態の表出方法
 - Dennettの意図・設計・物理スタンス

Human-Agent Interactionとは？ (5)

- 『HAIへの学際的アプローチ』 (小野 09 参照)
 - HAIの「危ない匂い」と「収まりの悪さ」
 - 細分化され、統合的な視点を欠いた研究への不満
 - 人間が命令すればそのとおり動く道具を作成することが目的ではない
 - 研究事例：
 - ITACO：メディア間を移動可能なエージェント
➡ 人工物にエージェント性を付与
 - Practical Magic：スマート情報環境とロボットのインタラクションデザイン
 - ITACO on the Room：部屋に移動するエージェント
 - RobotMeme：模倣と関係性の伝播
 - InteractionMeme：エージェント性と関係性の伝播によるインタラクション環境の構築
 - HAI研究はまさにムーブメントの結節点

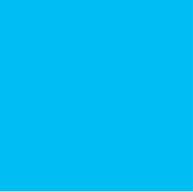
Human-Agent Interactionとは？ (6)

What is an agent? (国際会議HAIでの定義)

An agent is an object or technology that people interact with as if it is able to act with its own purposes, motivations, and intentions. Having autonomy is not sufficient. For example, people generally do not interact with telephones or elevators as agents although there is autonomy behind the scenes. Conversely, some people talk to their cars or boats as if they have their own intentions or personalities, even if this is not the case. It is the point where people shape their interactions with an object or technology as if it has purposes, motivations, or intentions, where we have human-agent interaction. Often, technologies are designed to encourage people to give them agency, for example, by using social methods such as speech or gestures, or having a personality, although this is not necessary for agency.

People attribute agency to many things and technologies in their everyday lives, which is a common area of research, for example, with robots, virtual characters and video games, virtual reality, smart cars, and so forth. By considering the broader field of Human-Agent Interaction, we have an opportunity to transcend these traditional categories and to share knowledge and learn from each other, regardless of the embodiment of our agent. Findings in human-robot interaction are likely to have implications for many kinds of agents.

The HAI conference aims to bring together an interdisciplinary community on agents. In addition to work on agents themselves, this conference also includes novel technologies and algorithms for use by agents, human-human studies that can inform interactive agent design, how psychological and cognitive models of agency are relevant to agents, and so forth.



エージェントを用いた対話システム

対話システムとは？ (1): 定型的な対話

- 人間と機械の間で、自然言語や音声言語を用いてコミュニケーションを行うことができるシステム
- 自然言語対話システム：
 - ELIZA(1965): 単純な構文解析、予め用意された表現

ユーザ： 男の人ってみんな同じなんだよね。

システム： どんなふうに？

ユーザ： 彼らはいつも何かで私たちを困らせるの。

システム： 具体的な例は思い付く？

ユーザ： そうね、私のボーイフレンドは私をここに来させたわ。

システム： あなたのボーイフレンドがあなたをここへ来させた。

ユーザ： 彼は私がいつも落ち込んでいるっていうの。

システム： あなたがいつも落ち込んでいるとは、お気の毒です。

(Weizenbaum, 1965)

対話システムとは？ (2): 意味理解

ー 自然言語対話システム:

- SHRDLU(1972): 限られた世界で、言語の理解を行うシステム ⇒ 取り扱う世界(ドメイン)で必要な知識をすべて書き込んでおく必要あり ⇒ フレーム問題

ユーザ: 大きい赤のブロックをつかんでください。

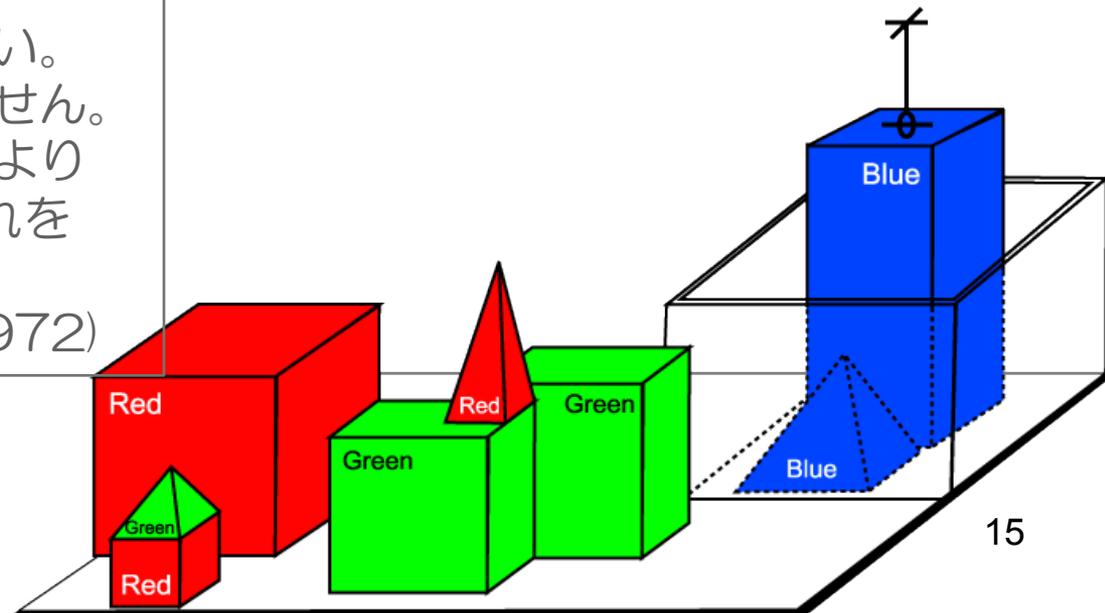
システム: はい、実行しました。

ユーザ: 四角錐をつかんでください。

システム: どの四角錐かわかりません。

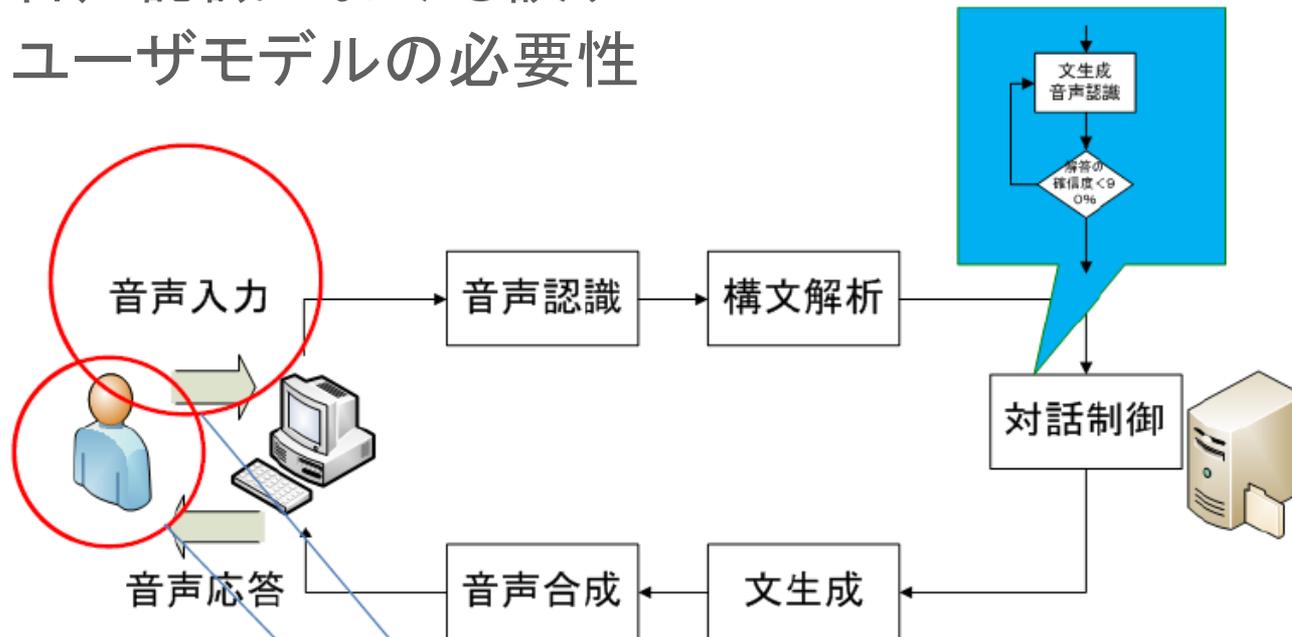
ユーザ: 今つかんでいるブロックより高いブロックを見つけて、それを箱の中に入れてください。

(Winograd, 1972)



対話システムとは？ (3): 音声対話

- 音声対話システム: 「音声」を用いる問題点
 - 言い誤りや倒置など文法的多様性
 - 音声認識における誤り
 - ユーザーモデルの必要性



- 音声認識誤りがある.
- ユーザーのモデル化が必要.

対話システムとは？ (4): 対話の主導権

- 対話システムの主導権：
 - システムが持つか、ユーザが持つか

システム主導対話

システム： ホテル検索・予約システムです。
宿泊先の最寄り駅をお話してください。

ユーザ： 上野駅。

システム： 上野駅周辺には15件の予約可能なホテルがあります。予算はいくらですか？

ユーザ： 8000円以内で。

システム： 何名様でしょうか。

ユーザ： 1人です。

システム： ABCホテル、ホテルDの2件見つかりました。

ユーザ： ABCホテルでお願いします。

ユーザ主導対話

ユーザ： 明後日から1泊、上野駅にホテルを予約したいんだけど。

システム： 上野駅周辺には15件の予約可能なホテルがあります。

ユーザ： じゃあ、8000円以内で泊まれるところは？

システム： ABCホテル、ホテルDの2件見つかりました。

ユーザ： ABCホテルにシングル1室予約して。

システム： ABCホテル、明後日から1泊シングルを1部屋予約しました。

音声認識率向上、意味理解容易

音声誤認識増加、意味理解難しい

対話システムとは？ (5): 混合主導権

－ 対話の主導と対話制御：

混合主導対話

ユーザ： 明後日から上野駅にホテルを予約してください。

システム： 何泊のご予定ですか？

ユーザ： 1泊。

システム： 予算はどれくらいでしょうか？

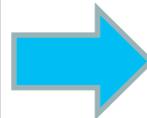
ユーザ： 8000円以内で。

システム： 何名様でしょうか？

ユーザ： 1名です。

システム： ABCホテル、ホテルDの2件見つかりました。

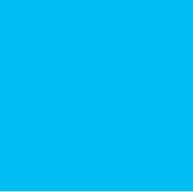
ユーザ： ABCホテルをお願いします。



・ 対話制御：

1. ユーザからの入力
2. ユーザの意図を検出
3. システムの適切な行動生成
4. 返すべき応答内容を生成

ユーザ主導 ⇒ システム主導

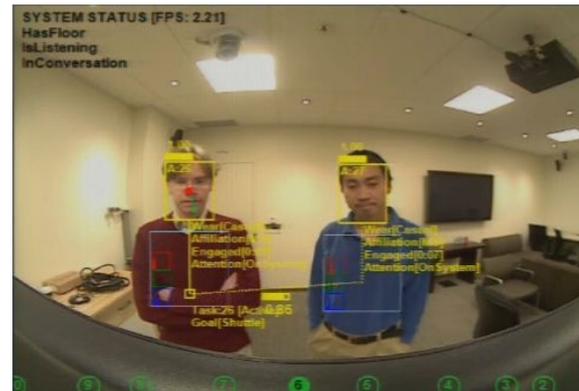


Situated Interaction

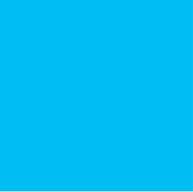
(MSR 2009, 2014)

近年の音声対話システム

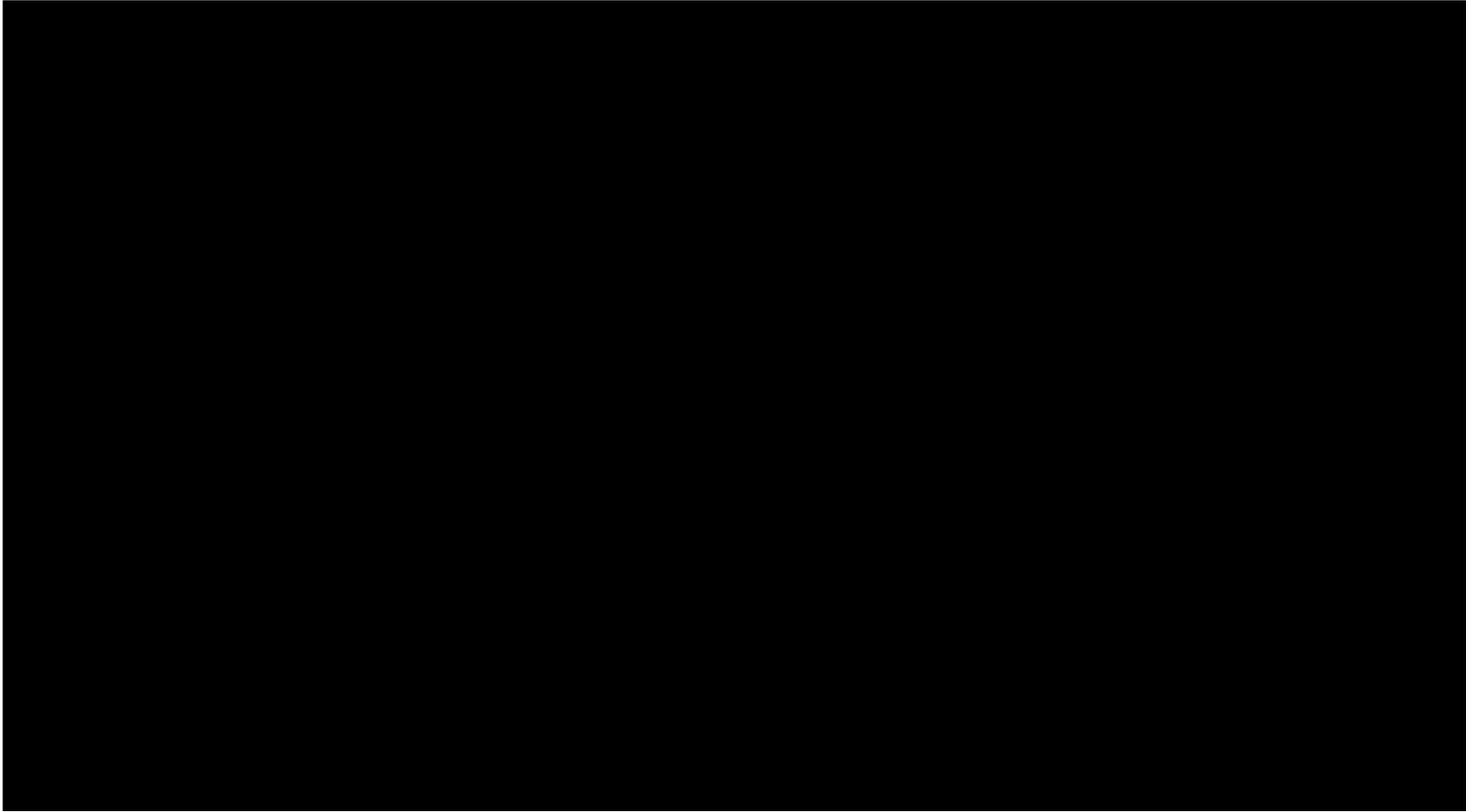
- 対話状況認識(会話シーン分析、マルチモーダルセンシング・フュージョン)
- 自然言語による多人数対話
- 対話制御(部分観測マルコフ決定過程)



Situated Interaction project (MSR)

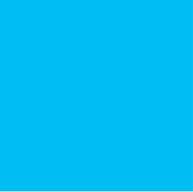


Situated Interaction (MSR 2009)



Situated Interaction (MSR 2014)





Mentalizing

**Rational quantitative attribution of beliefs,
desires and percepts in human mentalizing**
(Nature human behavior 2017)

概要

Social cognition depends on our capacity for ‘**mentalizing**’, or explaining an agent’s behaviour in terms of their mental states.

The development and neural substrates of mentalizing are well-studied, but its **computational basis** is only beginning to be probed. Here we present a model of core mentalizing computations: inferring jointly an actor’s beliefs, desires and percepts from how they move in the local spatial environment. Our **Bayesian theory of mind (BToM) model** is based on probabilistically inverting artificial-intelligence approaches to rational planning and state estimation, ...

BToM formalizes this schema as a **generative model** for action based on solving a **partially observable Markov decision process**, and formalizes mentalizing as **Bayesian inference** about unobserved variables (beliefs, desires, percepts) in this generative model, conditioned on observed actions.

HAI (HRI) の本質的な課題

人間の社会的知性



心の理論 = 他者の意図、願望、信念を推定



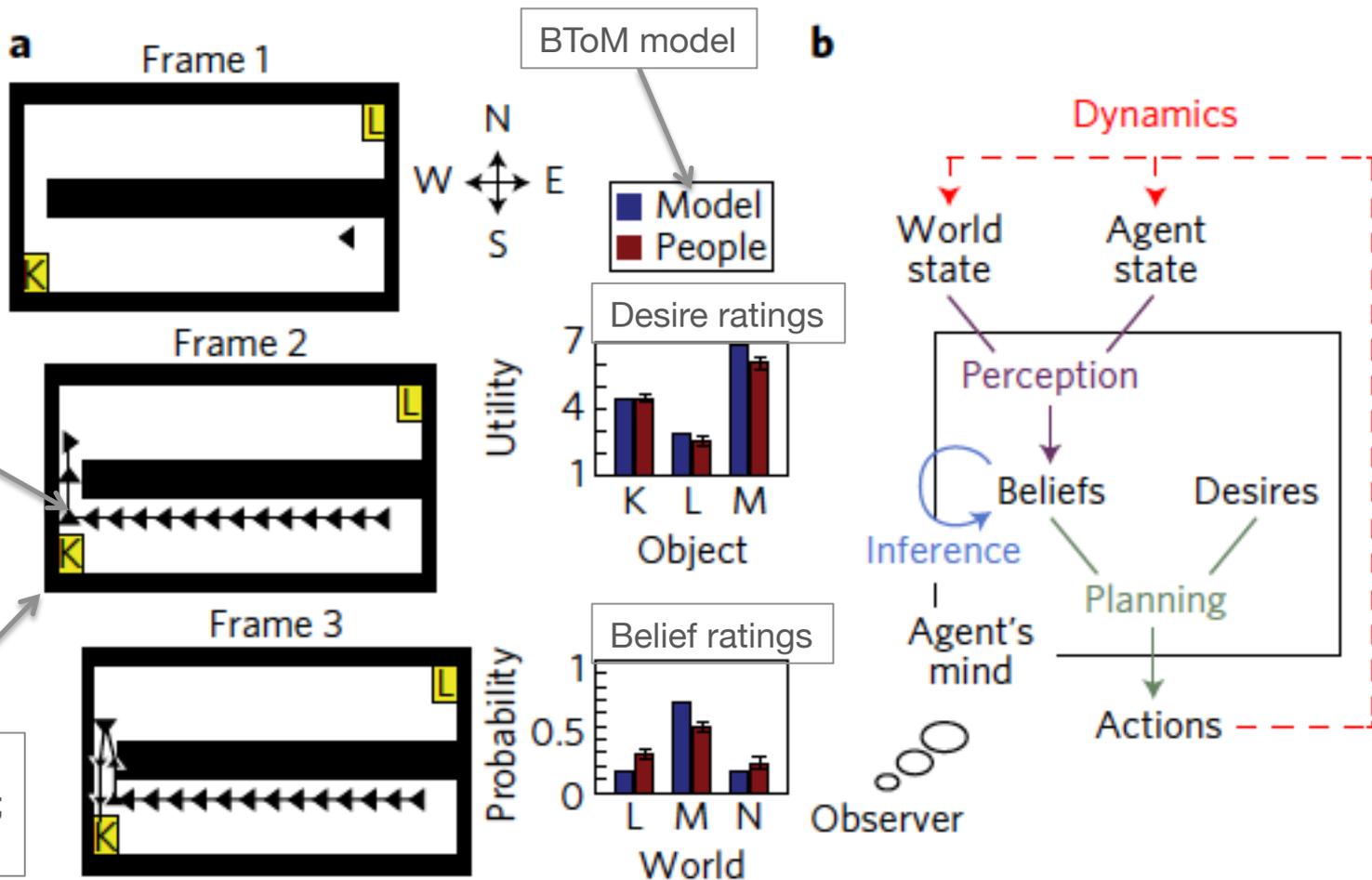
(部分情報から推定)



POMDP, Bayesian inference,
Generative model

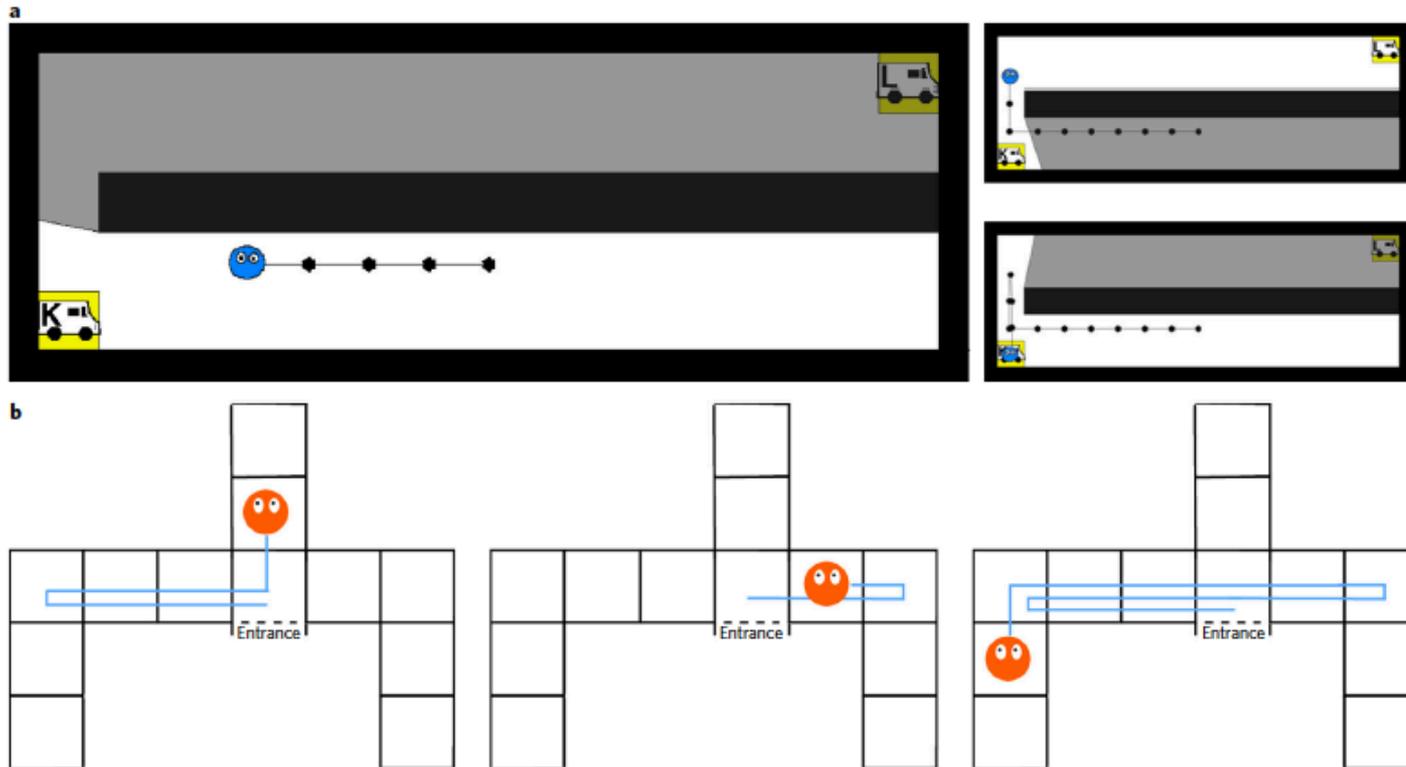
Mentalizing (1)

– 実験シナリオとモデル



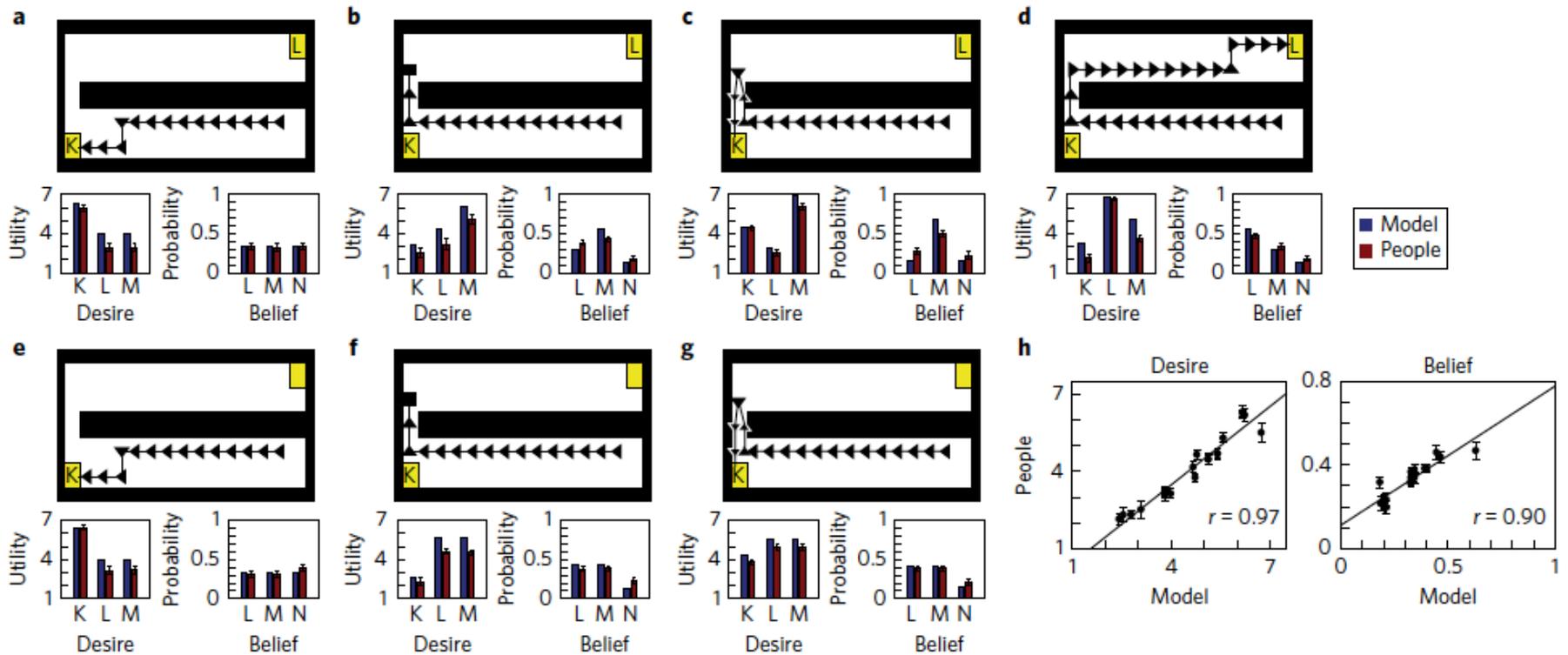
Hmentalizing (2)

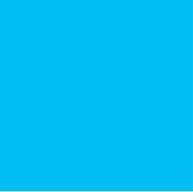
一 実験刺激の事例



Mentalizing (3)

— 実験1の結果





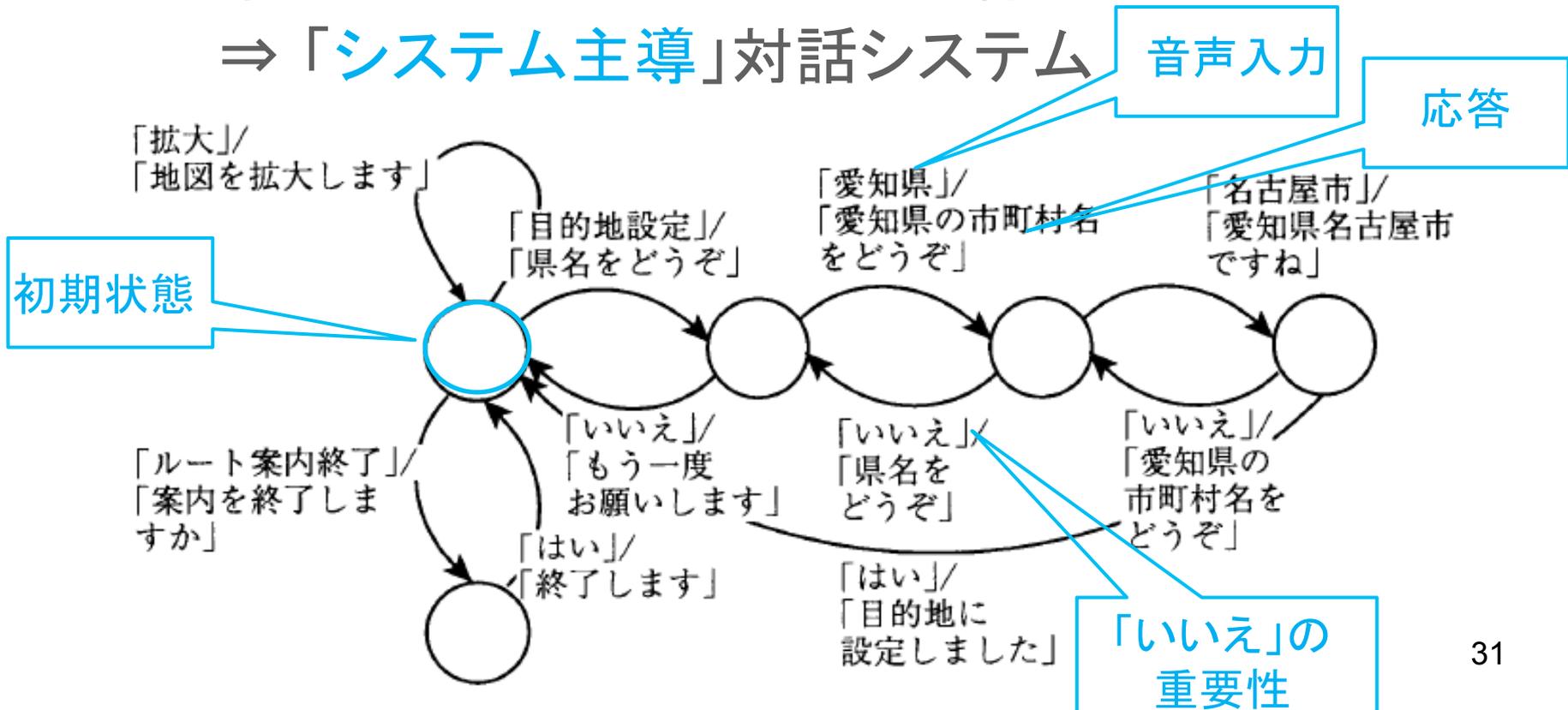
POMDP (Partially Observable Markov Decision Process) による対話制御

POMDPの背景

- マルコフ決定過程 ← 状態をすべて観測できる
- 部分観測マルコフ決定過程 ← 状態を部分的にしか観測できない
 - 「観測」と「状態」を分ける → 「信念」の確率分布で表現
 - 最終的に「状態価値関数」を推定したい
 - 正確に計算しようとするると計算量は指数関数的爆発 → 特定の点だけを求める方法がある

対話制御 (1): 有限状態オートマトン

- 有限状態オートマトンを用いた対話: 例. カーナビ
 - 動作: 初期状態 ⇒ 音声入力 ⇒ 応答 ⇒ 状態遷移
- 各状態におけるユーザの発話を予想
⇒ 「システム主導」対話システム



対話制御 (2): 意味表現

- ユーザ主導対話 ← ユーザの発話予想が難しい
 - 動作・処理:
 1. 音声認識・構文解析
 2. 意味解析と意味表現
 3. 文脈理解
 4. 応答生成

ユーザ： 河口湖にはどんなホテルがありますか？

システム： 河口湖ホテルやレイグランドホテル、
レジーナ河口湖があります。

ユーザ： テニスのできるホテルはありますか？

システム： レジーナ河口湖があります。

ユーザ： そのホテルの宿泊料金はいくらですか？

(伊藤、1998)

対話制御 (3): POMDP[1]

- 部分観測マルコフ決定過程 (partially observable Markov decision process, POMDP)
 - システムの状態を離散的に表現するのではなく、対話相手(ユーザ)が現在どの状態にあるかを確率的に表し、その確率の組みで連続的に表現する
- なぜ「部分観測」か？
 - POMDPで用いる「状態 S 」は、「ユーザの意図」、「ユーザの発話」、「対話履歴」の組みで表現
 - しかし、それらはシステムからは直接観測できず、観測できるのは、「ユーザ発話の認識結果 O 」のみ
- ユーザの状態を表現するPOMDP ← システムは一部の情報のみ観測可能 ⇒ 「部分観測」

対話制御 (3): POMDP[2]

- 例. ボイスメールを「保存」するか、「削除」するかを音声対話により指示する場合の、ユーザの「意図」と「発声」を状態として表現
- 削除/保存それぞれの状態 s にいる確率
⇒ 確率の組みがシステムの信念 (belief) b

「保存」状態

(ユーザが「保存」を意図し
「保存」と発声した状態)

$b = (1.0, 0.0)$

$b = (0.0, 1.0)$



「削除」状態

(ユーザが「削除」を意図し
「削除」と発声した状態)

対話制御 (3): POMDP[3]

- システムは信念 b に応じて応答 $a \in A$
- ユーザ発話の認識結果 $o \in O$
(ダッシュ記号は対話の1ターン後の変数)
- 状態遷移確率 $p(s' | a, s)$ はここでは変化しない

$$b'(s') = p(s' \circ o', a, b)$$

システムの定義より

$$= \frac{p(o' \circ s', a, b) p(s' | a, b)}{p(o' | a, b)}$$

ベイズ則による変形

確率の基本的な概念 (5): ベイズ則

- ベイズ則: 確率ロボティクスでは支配的な役割を果たす
- 条件付き確率 $p(x | y)$ をその「逆」 $p(y | x)$ と関連付ける ($p(y) > 0$)

$$p(x | y) = \frac{p(y|x)p(x)}{p(y)} = \frac{p(y|x)p(x)}{\sum_{x'} p(y|x')p(x')} \quad (\text{離散系})$$

$$p(x | y) = \frac{p(y|x)p(x)}{p(y)} = \frac{p(y|x)p(x)}{\int p(y|x')p(x')dx'} \quad (\text{連続系})$$

対話制御 (3): POMDP[4]

ロボット位置
の確率密度

$$b'(s') = p(s' | o', a, b)$$

システムの定義より

$$= \frac{p(o' | s', a, b) p(s' | a, b)}{p(o' | a, b)}$$

ベイズ則による変形

$$= \frac{p(o' | s', a) \sum_{s \in \mathcal{S}} p(s' | a, b, s) p(s | a, b)}{p(o' | a, b)}$$

全確率の定理

$$= \frac{p(o' | s', a) \sum_{s \in \mathcal{S}} p(s' | a, s) b(s)}{p(o' | a, b)}$$

$p(s' | a, b, s) = p(s' | a, s)$ と仮定
※確率的発生法則(1)参照

$$= \eta p(o' | s', a) \sum_{s \in \mathcal{S}} p(s' | a, s) b(s)$$

分母は s' に無関係のため η

計測モデル

動作モデル

t-1のロボット位置の確率密度関数

ベイズフィルタアルゴリズム参照

確率の基本的な概念 (4)

- 全確率の定理: 条件付き確率の定義と確率測度の公理から導かれる ($p(y) = 0$ であればゼロ)

$$p(x) = \sum_y p(x|y)p(y) \quad (\text{離散系の場合})$$

$$p(x) = \int p(x|y)p(y)dy \quad (\text{連続系の場合})$$

確率的発生法則 (1)

- 状態 x が完備、制御動作 u の後に計測 z を行うとき

$$p(x_t | x_{0:t-1}, z_{1:t-1}, u_{1:t}) = p(x_t | x_{t-1}, u_t)$$

完備性ゆえ
 u_t だけを考慮

- 上式は、条件付き独立性の例。計測も同様に

$$p(z_t | x_{0:t}, z_{1:t-1}, u_{1:t}) = p(z_t | x_t)$$

- つまり、ばらつきのあり得る計測値 z_t を予測するためには x_t だけわかればよい
- x_t が完備であれば、過去の計測、制御、状態は関係ない

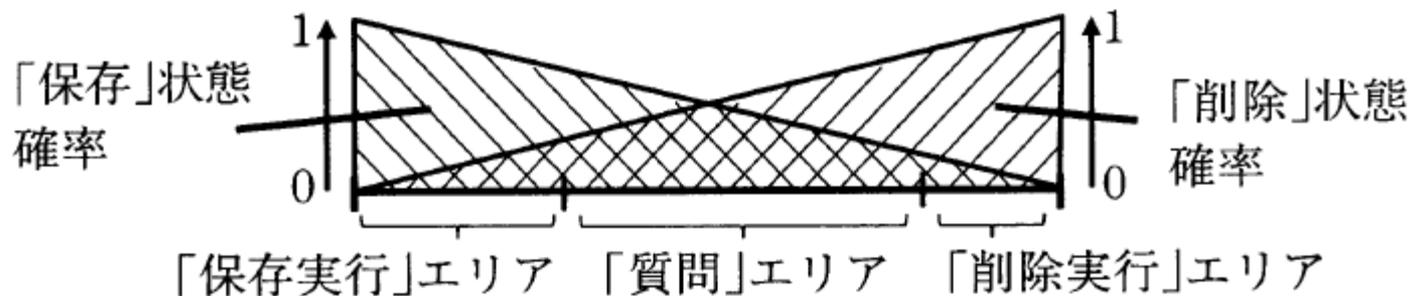
条件付き独立性は重要



アルゴリズムが計算機と相性がよくなる

対話制御 (3): POMDP[5]

- ユーザの発話の誤認識:
 $p(o' = delete | s' = delete, a = question) = 0.7$
 $p(o' = save | s' = save, a = question) = 0.8$
- システムの応答 a は信念 b を政策 π によってマッピング $a = \pi(b)$
- 報酬 $r(s, a)$



「政策」の分割

対話制御 (3): POMDP[6]

ユーザとシステムの対話例:

0. 初期信念:

$$b = (b(\text{save}), b(\text{delete})) = (0.65, 0.35)$$

1. 初期信念は「質問」エリアにある

⇒ システムは質問「save or delete?」、ユーザは「save」と発話

2. システムは「delete」と誤認識

3. 「信念」の更新 ⇒ 「delete」よりの信念に

4. まだ「質問」エリアのため再度質問 ⇒ システムは「save」と正しく認識(2回)

5. 「信念」の更新 ⇒ 「保存実行」エリアへ ⇒ 「保存」

対話制御 (3): POMDP[7]

システムの得る報酬: $r(s, a)$

- 正しい動作(応答):

$$r(\text{save}, \text{save}) = 5$$

$$r(\text{delete}, \text{delete}) = 5$$

- 間違った動作:

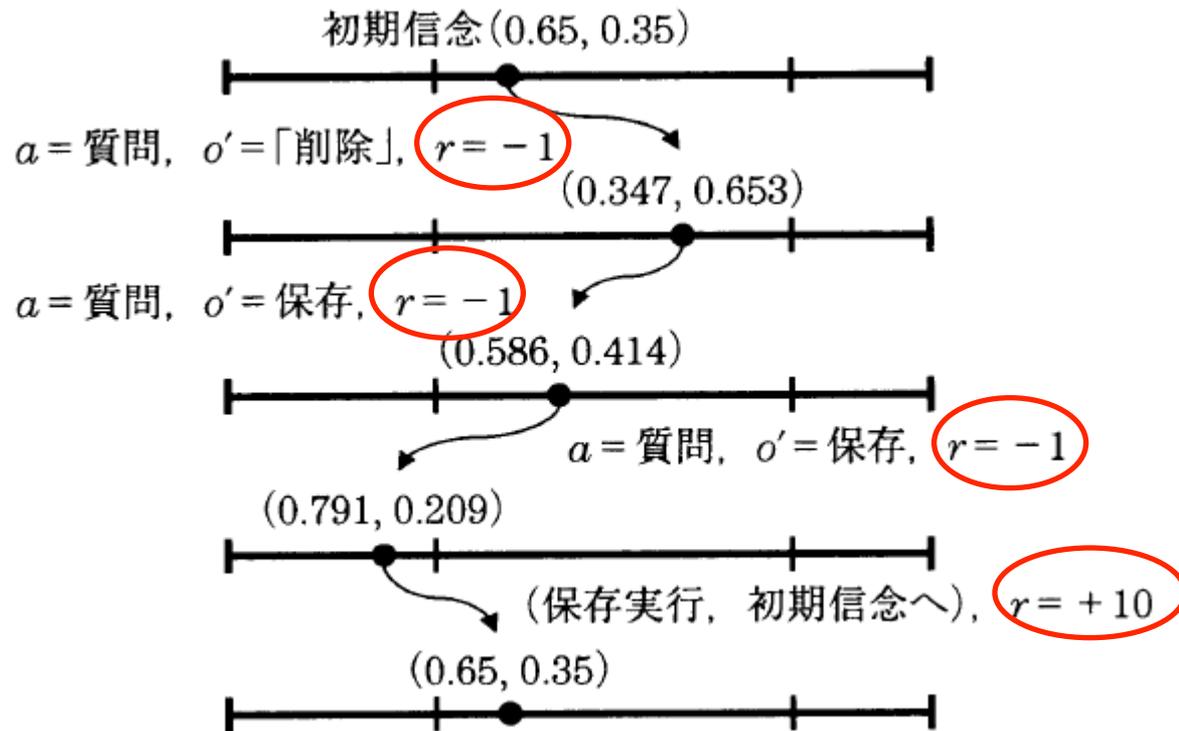
$$r(\text{save}, \text{delete}) = -20$$

$$r(\text{delete}, \text{save}) = -10$$

- 「質問」:

$$r(\text{save}, \text{question}) = -1$$

$$r(\text{delete}, \text{question}) = -1$$



報酬合計 $r = 7$

例題

- スライド「対話制御(3):POMDP[7]」のように、本システムでは誤認識により、信念が以下のように更新されている。本講義で用いた確率値と、スライド「対話制御(3):POMDP[4]」の式を用いてこの更新を確認せよ。
 - 初期信念:
 $b = (b(\text{save}), b(\text{delete})) = (0.65, 0.35)$
 - 更新された信念:
 $b = (b(\text{save}), b(\text{delete})) = (0.347, 0.653)$

例題の解答

更新式に確率を代入して

$$\begin{aligned} b'(save) &= \eta p("del" | save, que) \{p(save | que, save)b(save) + \\ & p(save | que, del)b(del)\} \\ &= \eta \cdot 0.2 \cdot \{1.0 \cdot 0.65 + 0.0 \cdot 0.35\} = \eta \cdot 0.130 \end{aligned}$$

$$\begin{aligned} b'(del) &= \eta p("del" | del, que) \{p(del | que, save)b(save) + \\ & p(del | que, del)b(del)\} \\ &= \eta \cdot 0.7 \cdot \{0.0 \cdot 0.65 + 1.0 \cdot 0.35\} = \eta \cdot 0.245 \end{aligned}$$

η は正規化係数なので

$$\begin{aligned} \eta(0.130 + 0.245) &= 1 \\ \eta &= 2.667 \end{aligned}$$

ゆえに

$$b'(0.347, 0.653)$$

対話制御 (3): POMDP[8]

- 強化学習による最適化 (Value Iteration法):
 - システムが将来獲得できる報酬を最大化する a を実行する政策を学習する
 - システムが将来の時刻 t までに得る報酬を信念の価値関数 $V_t(b)$ で表し、 $t \rightarrow \infty$ とすると

$$V_{\infty}(b) = \sum_{\tau=0}^{\infty} \gamma^{\tau} \sum_s b_{\tau}(s) r(s, a_{\tau})$$

が将来得る総報酬となる

ここで、割引率 γ であり、近い将来ほど影響大 ($0 \leq \gamma \leq 1$)

対話制御 (3): POMDP[9]

- これは、次式と、以下の漸化式を用いて得られる。

$$V_1(b) = \max_a \sum_s b(s)r(s, a)$$

$$V_{T+1}(b) = \max_a \left\{ \sum_s b(s)r(s, a) + \gamma \sum_{o'} P(o' | a, b) V_T(b', (s')) \right\}$$

対話制御 (3): POMDP[10]

- 部分観測マルコフ決定過程 (partially observable Markov decision process, POMDP)
 - システムの状態を離散的に表現するのではなく、対話相手(ユーザ)が現在どの状態にあるかを確率的に表し、その確率の組みで連続的に表現する
- なぜ「部分観測」か？
 - POMDPで用いる「状態 s 」は、「ユーザの意図」、「ユーザの発話」、「対話履歴」の組みで表現
 - しかし、それらはシステムからは直接観測できず、観測できるのは、「ユーザ発話の認識結果 o 」のみ
- ユーザの状態を表現するPOMDP←システムは一部の情報のみ観測可能 ⇒ 「部分観測」

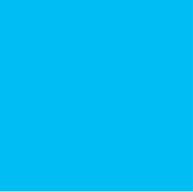
参考文献：対話システム

[Reference]

- J. Williams, "Statistical approaches to dialogue systems," Tutorial at *Interspeech* (2009).
- J. Williams, P. Poupart and S. Young, "Partially observable Markov decision processes with continuous observations for dialogue management," *SIGdial*, pp. 25-34 (2005).
- D. Bohus, E. Kamar and E. Horvitz, "Towards Situated Collaboration," in *NAACL Workshop on Future Directions and Challenges in Spoken Dialog Systems: Tools and Data* (2012).
- D. Bohus, E. Horvitz, "On the Challenges and Opportunities of Physically Situated Dialog," in *AAAI Fall Symposium on Dialog with Robots* (2010).

[HP] Situated Interaction Project in MSR:

http://research.microsoft.com/en-us/um/people/dbohus/research_situated_interaction.html



Human-Agent Interaction (HAI) (Human-Robot Interaction (HRI))

なぜ対話エージェントなのか？ (1)

– 最近の動向 . . .

- 2016年4月 FacebookがMessenger上での対話型エージェントの機能を発表
- 2016年9月 LINEがボット (Bot) 向けの「Messaging API」を発表
- 2014年 Amazonが音声対話デバイス「Amazon Echo」を発表
- 2016年 Googleが音声対話デバイス「Google Home」を発表
- (Appleは「HomePod」の発売延期を発表 . . .)

– しかし、問題も . . .

- 2016年 Microsoftのチャットロボット「Tay」が差別発現やナチスを賛美する発言を繰り返し、16時間で停止



なぜ対話エージェントなのか？ (2)

- 工学的には . . .
 - GUIとは違い、直感的に使うことができる
 - 「友だち」を追加する感覚で導入できる
 - キャラクター(人格)の存在による親しみやすさ(萌え家電)
- ➡ マルチモーダルコミュニケーション
- 科学的には . . .
 - 人間の擬人化する能力：Media Equation、アニミズムなど
 - ミラーニューロン(mirror neuron)：他の個体の行動を見て、まるで自身が同じ行動をとっているかのように脳が賦活
 - 心の理論(theory of mind)：他者の心を読む、脳のモジュール構造、他者モデル
 - ノンバーバルコミュニケーション（ジェスチャー）：考える身体

なぜ対話エージェントなのか？ (3)

対話型エージェントの分類：タスク型と非タスク型

用途	概要	例	導入の特徴
タスク	非 スタンド アロン型	カスタマー サポート型	<p>ユーザーからの問い合わせを対話UIで実現するもの。ユーザーが慣れ親しんだUIでキャラクターに質問をさせることで、満足度向上/エンゲージメント向上につながる</p> <p>LOHACO 「マナミさん」</p> <ul style="list-style-type: none"> 業界業種を問わず導入することで効果を上げられる 大規模な開発は必要ないため導入の敷居が低く、短期のローンチが可能
		プッシュ型	<p>対話UIが双方向であるという特性を生かし、プッシュ型で情報を届けることを主な目的とするもの</p> <p>天気予報、 ニュース</p> <ul style="list-style-type: none"> メディア系のサービスが中心 自社サービスと接続されたシステム開発が可能
		決済 チャンネル型	<p>商品やサービスの予約・購入を対話UI上で行えるもの。アカウントがひも付いており、アクセスが容易であるという特性を生かし、販売チャンネルを広げることが可能</p> <p>航空券予約</p> <ul style="list-style-type: none"> ECサイトが中心 自社サービスと接続されたシステム開発が可能
	スタンド アロン型	<p>対話UIだからこそ実現できるような、特定の目的を達成するためのサービス</p> <p>アップル 「Siri」、 x.ai「x.ai」</p> <ul style="list-style-type: none"> 大手IT企業や個別のスタートアップが参入 	
非タスク		<p>会話そのものが目的であるサービス（雑談、会話型のゲームなど）</p> <p>マイクロソフト 「りんな」(日本) 「xiaoice」 (中国)</p> <ul style="list-style-type: none"> 大手IT企業や個別のスタートアップが参入 	

エージェントの
“ある” or “なし”
の効果はよくわからない？

エージェントの“ある” or “なし” (1)

- “Amazon Echo”



エージェントの“ある” or “なし” (2)

– VITA: Virtual Interactive Training Agent

- Virtual Interactive Training Agent (VITA) is a virtual reality job interview practice system for building competence and reducing anxiety in young adults with Autism Spectrum Disorder (ASD) and other developmental disabilities. It was developed by the USC Institute for Creative Technologies (ICT), in partnership with the Dan Marino Foundation (DMF).
- Combining DMF expertise on ASD and ICT virtual human technology, VITA provides the opportunity for ASD users to repetitively practice job interviewing in a safe simulated virtual reality environment. Various scenarios have been created that simulate real life job interview situations. Video recording of the user's interaction allows a vocational expert and the user to visually review and analyze the user's performance.
- Learn more at:
<http://www.ict.usc.edu/prototypes/vita/>

エージェントシーの“ある” or “なし” (3)

- “VITA” (overview)



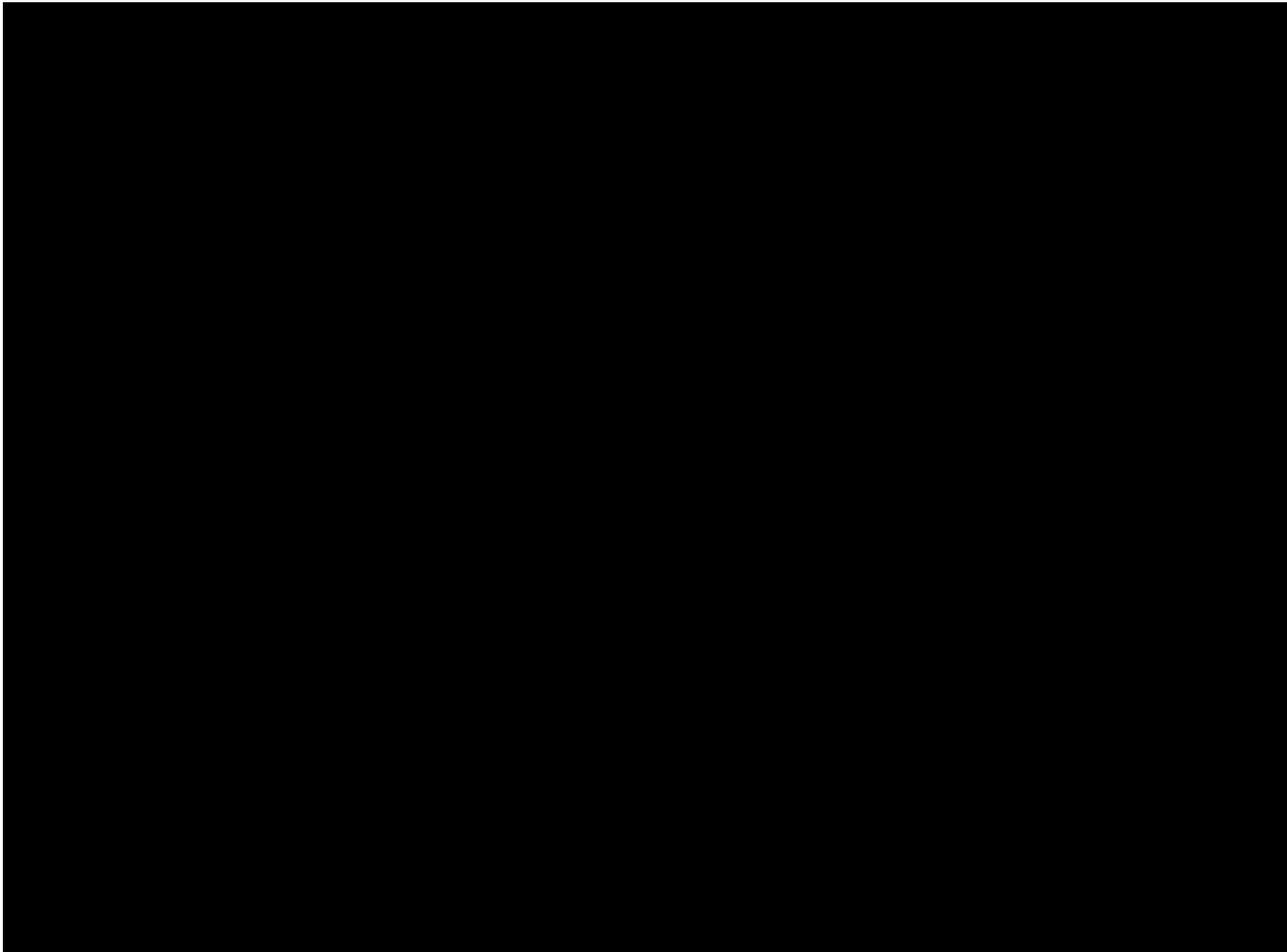
エージェントの“ある” or “なし” (4)

- “SimSensei” (cases of guidance)



エージェントシーの“ある” or “なし” (5)

- 自閉症児が積極的にコミュニケーションを行うロボット“キーポン”



エージェントシーの“ある” or “なし” (6)

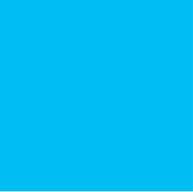
- エージェントシーの“ある” or “なし”を体感？
 - 就職の面接官はどちらが緊張しないか？

Amazon Echo or VITA or Keepon?

- 一緒にお酒を飲むとしたらどちらか？

Amazon Echo or VITA or Keepon?





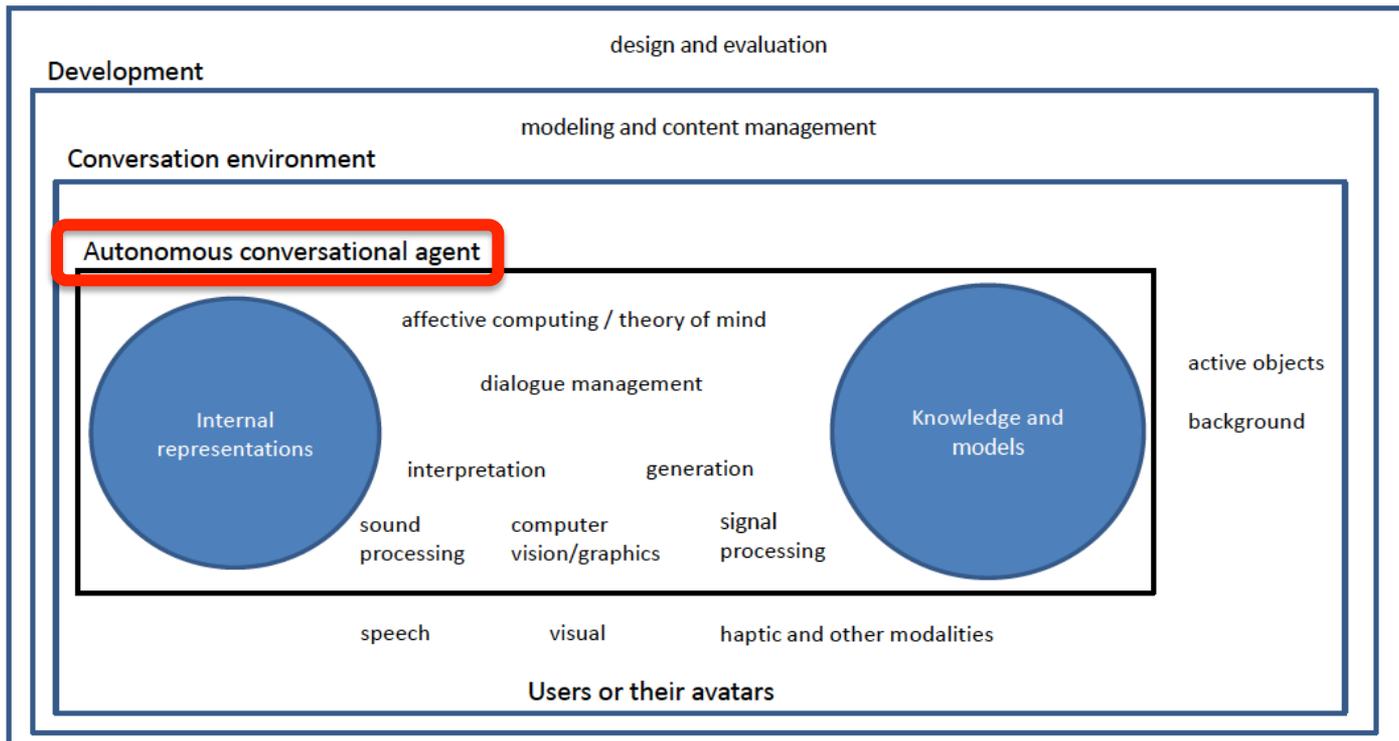
Embodied Conversational Agents (ECA) の構築方法

ECAの構築方法 (1)

ー 研究テーマ

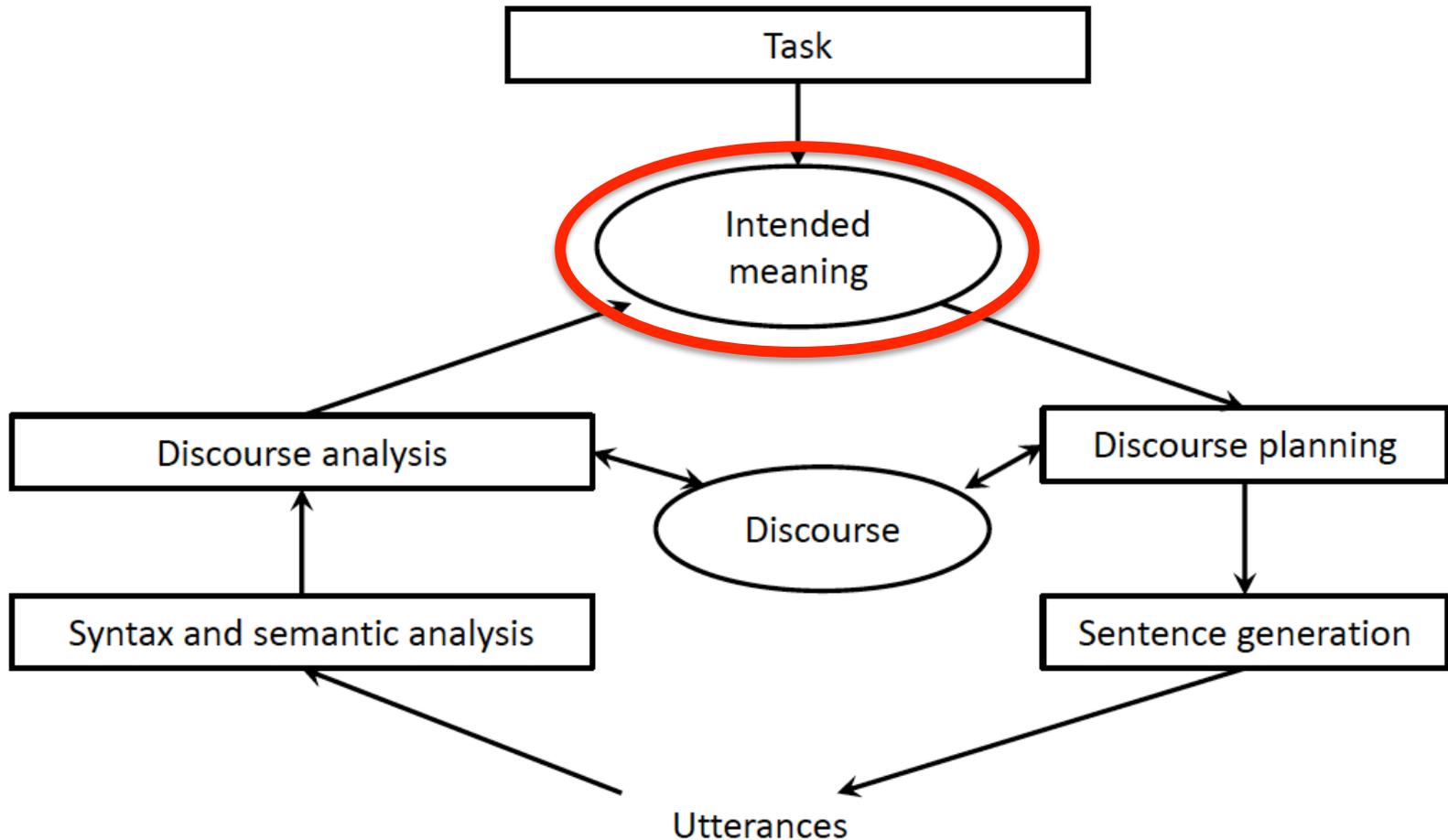
- ・ アーキテクチャ
- ・ スクリプトとマークアップ言語
- ・ コーパスベースド・アプローチ
- ・ 評価方法

Research



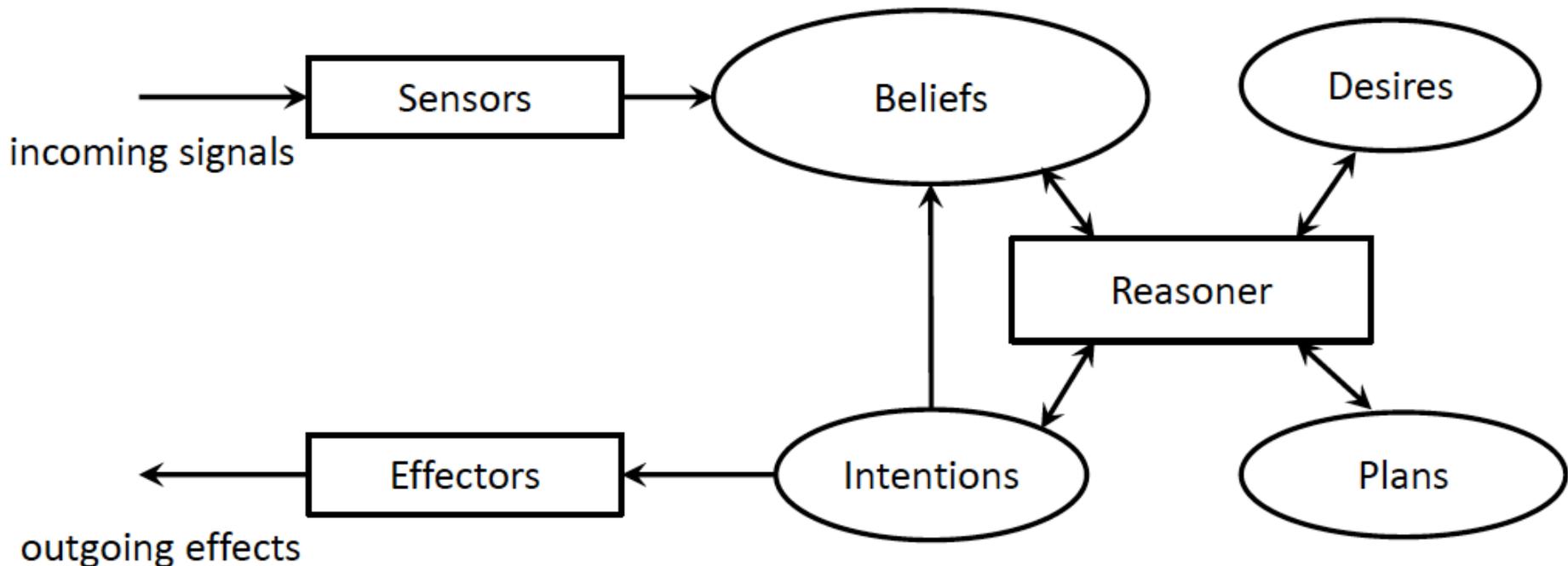
ECAの構築方法 (2)

– 対話の管理



ECAの構築方法 (3)

- 認知的コンピューティング：BDIアーキテクチャ

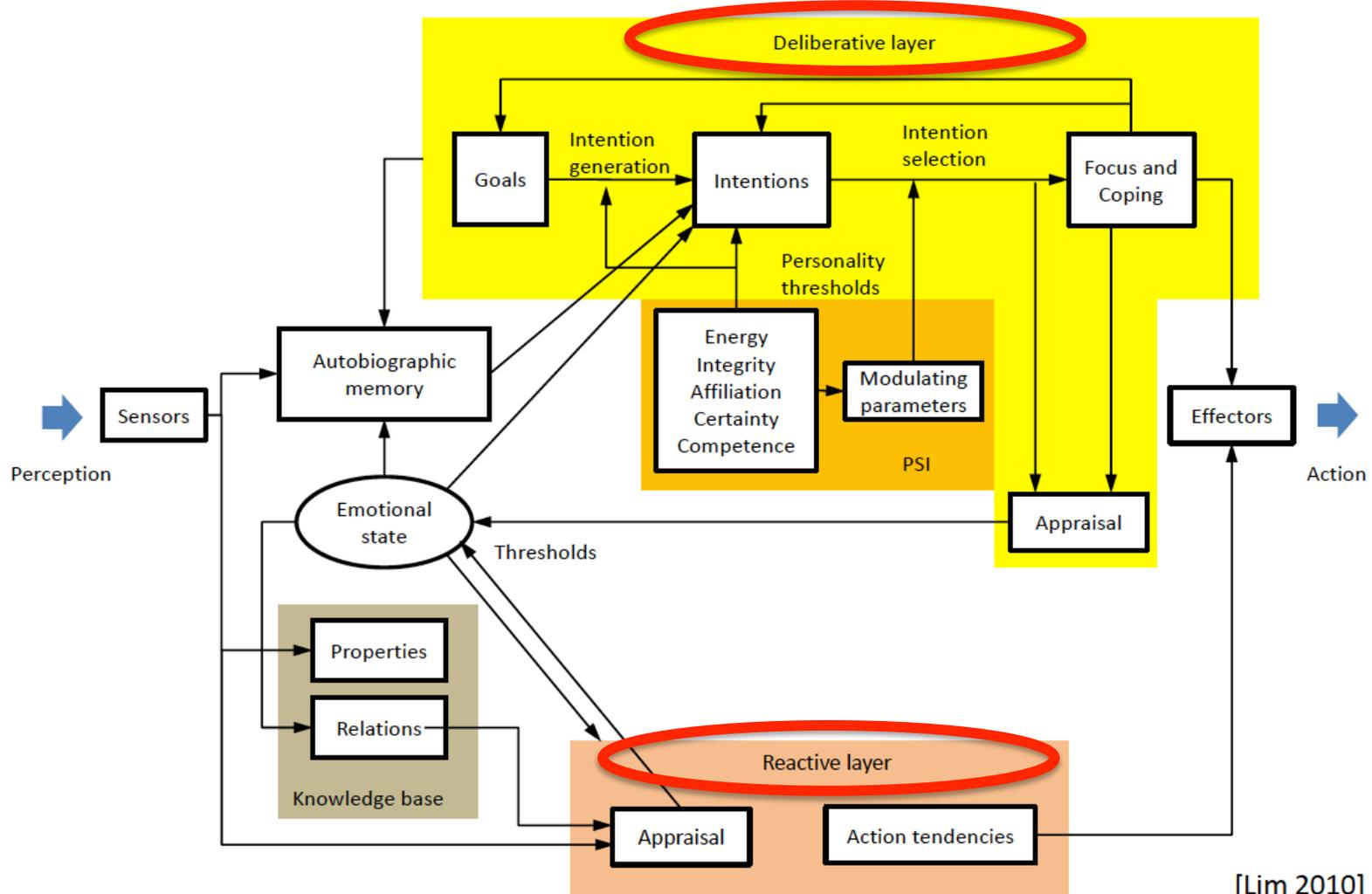


BDIアーキテクチャ：

自律エージェントの実現方法の1つであり、目標を達成するための行動の選択として「意図」を保持し、これに沿って行動することで一貫性を保った目標達成への行動が可能

ECAの構築方法 (4)

ー 認知的アーキテクチャ



ECAの構築方法 (5)

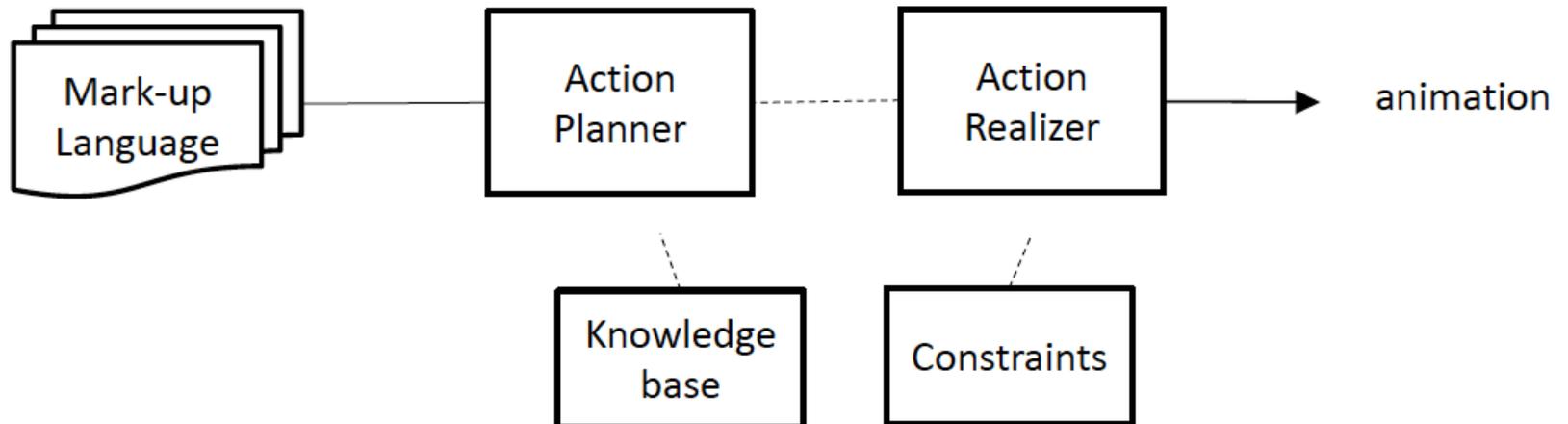
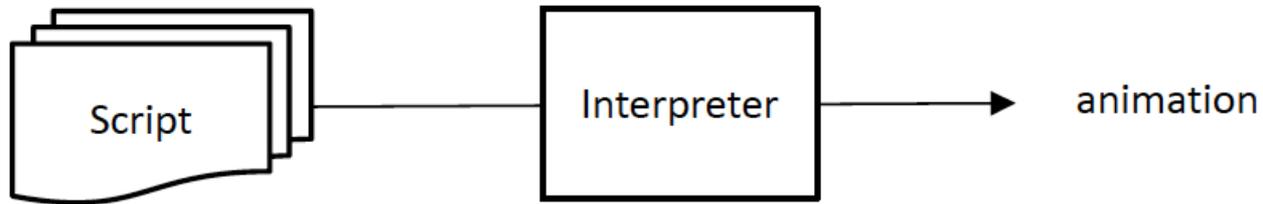
- スクリプト言語、マークアップ言語、Believable Agentの構成の必要性
 - ・ 同期することが可能な発話、視線、ジェスチャ
 - ・ ジェスチャ、顔表情の表出、発話などで性格や感情の表現
 - ・ 会話の中で、複数のキャラクターの身体動作の調整
 - ・ ユーザや他のエージェントとのコミュニケーション



システム環境からの支援が不可欠

ECAの構築方法 (6)

– Script vs Mark-up Language



ECAの構築方法 (7)

– Artificial Intelligence Mark-up Language (AIML)

C: Knock knock.
R: Who's there?
C: Banana.
R: Banana who?
C: Knock knock.
R: Who's there?
C: Banana.
R: Banana who?
C: Knock knock.
R: Who's there?
C: Orange.
R: Orange who?
C: Orange you glad I didn't say banana.
R: Ha ha very funny, Nancy.



```
<category>
<pattern>KNOCK KNOCK</pattern>
<template>Who is there?</template>
</category>

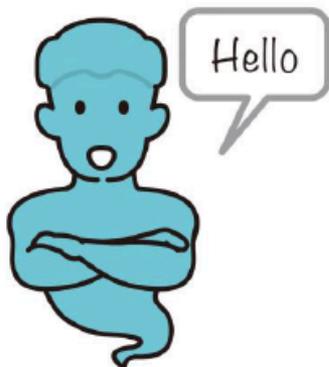
<category>
<pattern>*</pattern>
<that>WHO IS THERE</that>
<template><person/> who?</template>
</category>

<category>
<pattern>*</pattern>
<that>* WHO</that>
<template>Ha ha very funny, <get
name="name"/>.</template>
</category>
```

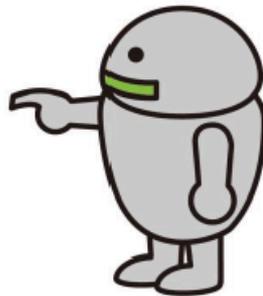
In AIML the syntax <that>...</that> encloses a pattern that matches the robot's previous utterance.

ECAの構築方法 (8)

– Microsoft Agent



Genie

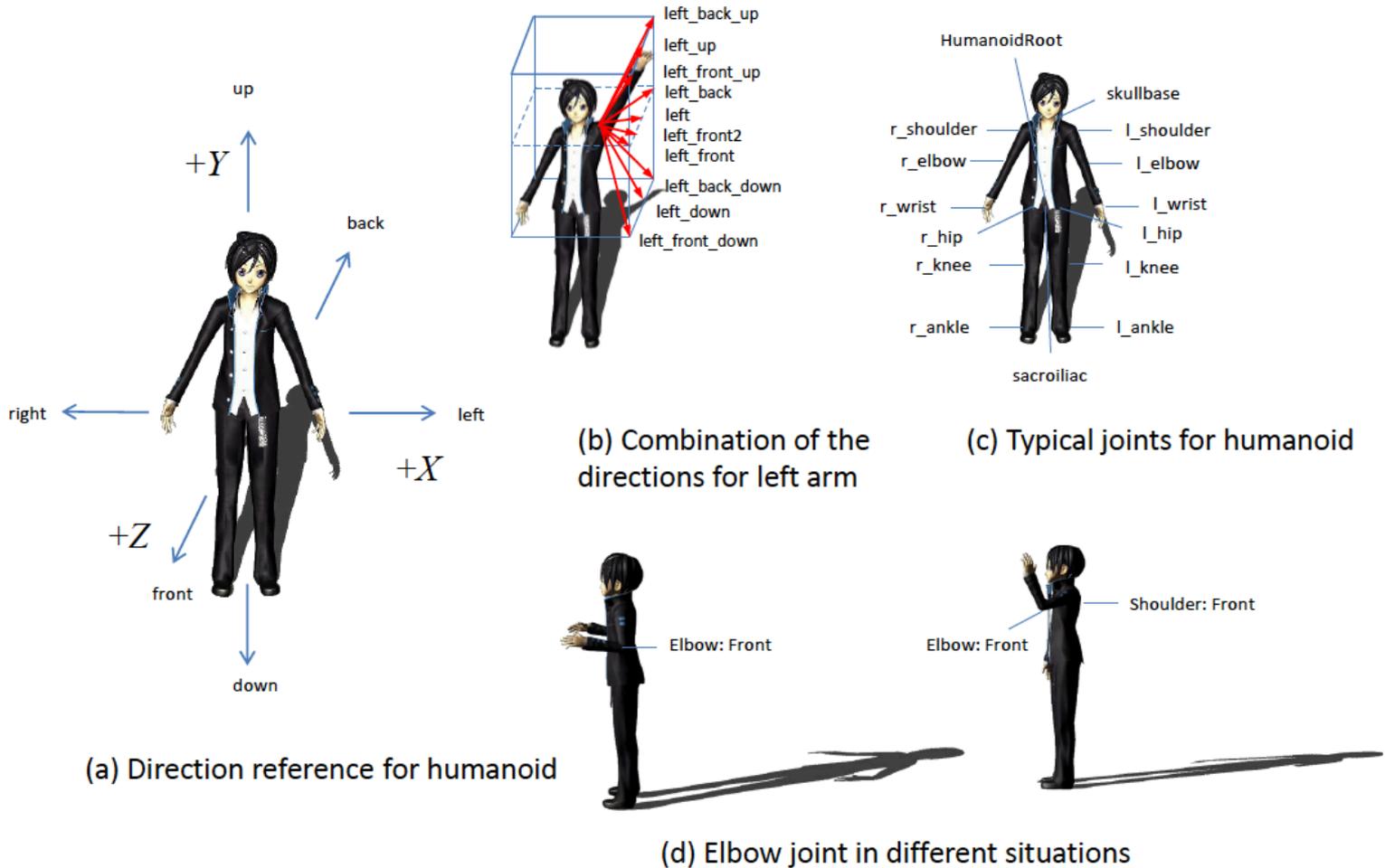


Robby

```
...  
Agent1.Characters.Load "Genie", "http: ..."  
Set Genie = Agent1.Characters("Genie")  
Set Robby = Agent1.Characters ("Robby")  
Genie.Show  
Genie.Play "Greet"  
Genie.Speak "Hello!"  
Robby.Show  
Robby.MoveTo 100,200  
Robby.Play "GestureRight"  
Robby.Hide  
Genie.Hide  
...
```

ECAの構築方法 (9)

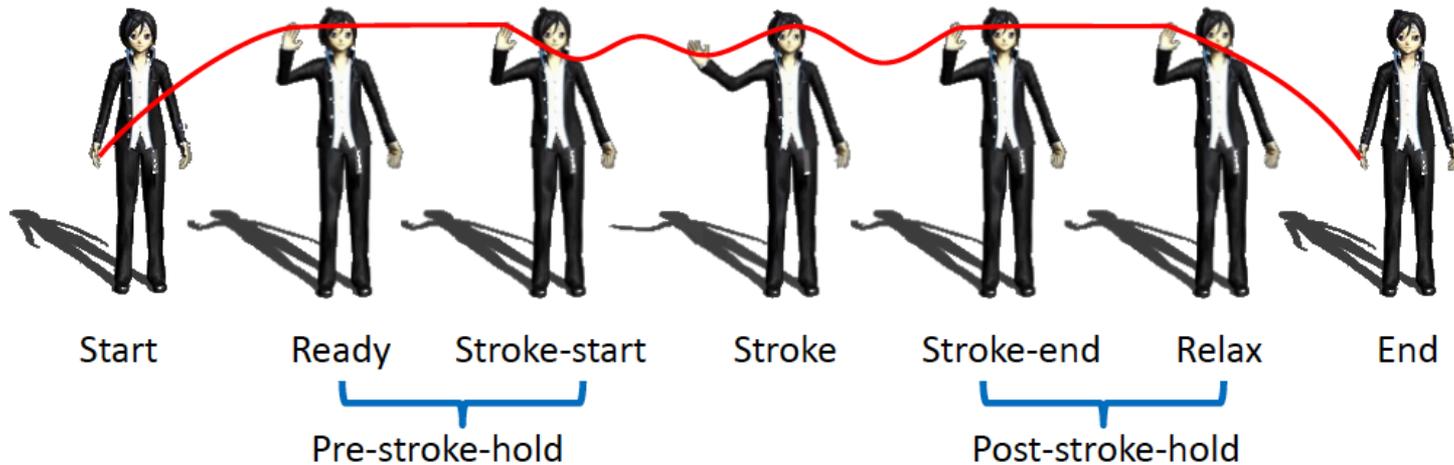
人型エージェントのパラメータ化



ECAの構築方法 (10)

- Behavior Markup Language (BML)

Synchronization



The seven synch points

ECAの構築方法 (11)

– BML behavior elements

BML Element	Description
<head>	Movement of the head independent of eyes. Type include nodding, shaking, tossing and orienting to a given angle.
<torso>	Movement of the orientation and shape of the spine and shoulder.
<face>	Movement of facial muscles to form certain expressions. Types include eyebrow, eyelid and larger expressive mouth movements.
<gaze>	Coordinated movement of the eyes, neck and head direction, indicating where the character is looking.
<body>	Full body movement, generally independent of the other behaviors. Types include overall orientation, position and posture.
<legs>	Movements of the body elements downward from the hip: pelvis, hip, legs including knee, toes and ankle.
<gesture>	Coordinated movement with arms and hands, including pointing, reaching, emphasizing (beating), depicting and signaling.
<speech>	Verbal and paraverbal behavior, including the words to be spoken (for example by a speech synthesizer), prosody information and special paralinguistic behaviors (for example filled pauses).
<lips>	This element is used for controlling lip shapes including the visualization of phonemes.

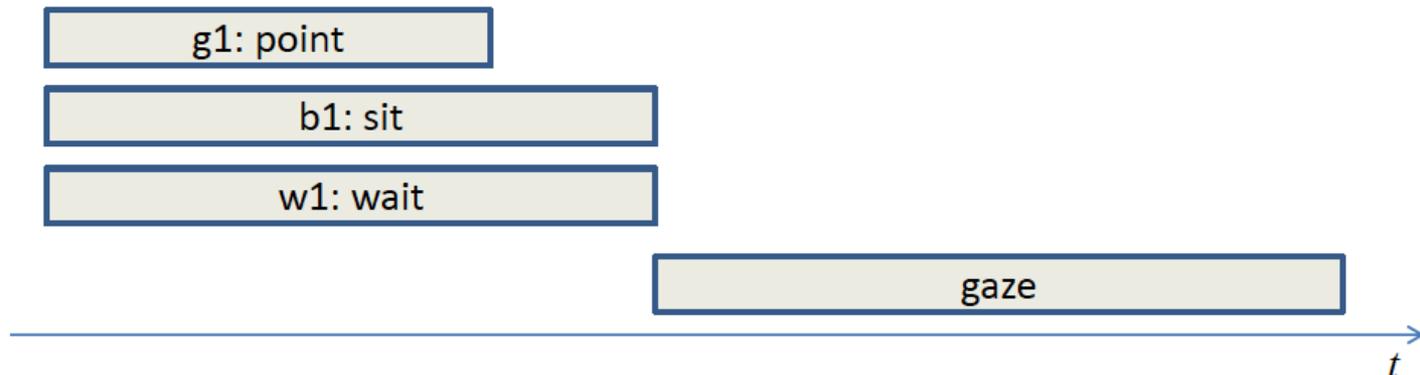
ECAの構築方法 (12)

– BML notation (1)

<wait> can align a behavior with a condition or an event

Example 1

```
<bml>  
  <gesture id="g1" type="point" target="object1"/>  
  <body id="b1" posture="sit"/>  
  <wait id="w1" condition="g1:end AND b1:end"/>  
  <gaze target="object2" start="w1:end"/>  
</bml>
```



ECAの構築方法 (13)

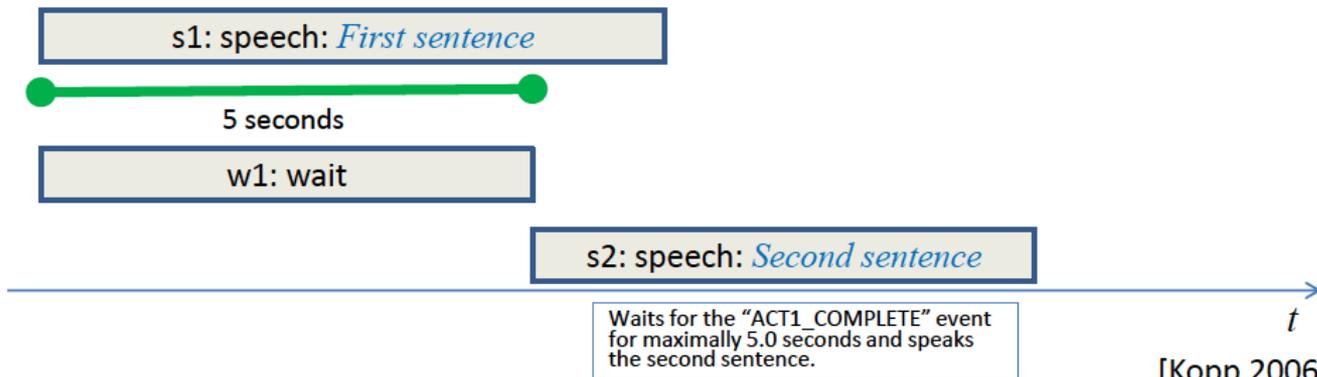
– BML notation (2)

<wait> can align a behavior with a condition or an event

Example 2

```
<bml>
  <speech id="s1" type="text/plain">
    First sentence
  </speech>
  <event start="s1:end" emit="ACT1_COMPLETE" />
</bml>
```

```
<bml>
  <wait id="w1" event="ACT_COMPLETE" duration="5.0"/>
  <speech type="text/plain" start="w1:end">
    Second sentence
  </speech>
</bml>
```



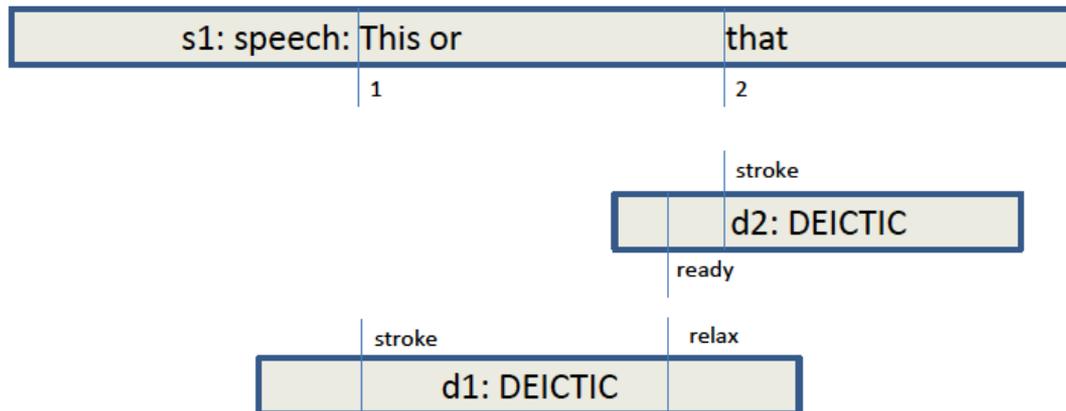
ECAの構築方法 (14)

– BML notation (3)

Timing holds

Example 3

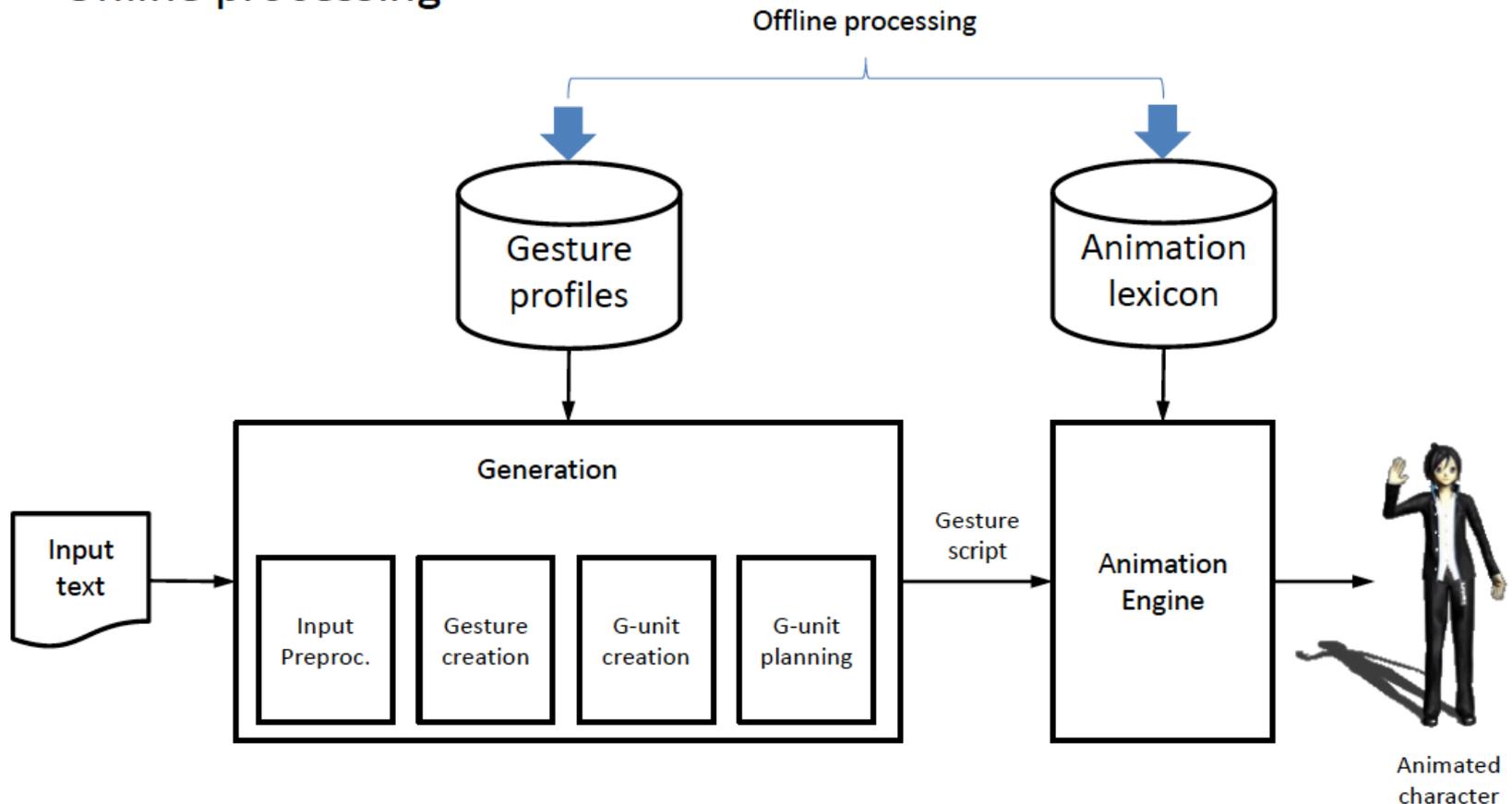
```
<speech id="s1"> <sync id="1"/>This or <sync id="2"/>that.</speech>  
...  
<gesture id="d2" type="DEICTIC" ... stroke="s1:2" />  
...  
<gesture id="d1" type="DEICTIC" ... stroke="s1:1" relax="d2:ready" />
```



ECAの構築方法 (15)

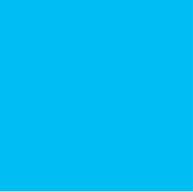
– Generating natural gestures

Online processing



ECAの構築方法 (16)

- ECAの評価方法
 - ユーザビリティ
 - 学習可能性 (learnability)、記憶しやすさ (memorability)、使い勝手の良さ (ease of use)
 - 効率性 (efficiency)
 - エラー発生頻度 (frequency of errors)
 - ユーザ知覚
 - 満足度 (satisfaction)、関与 (engagement)、有用性 (helpfulness)、自然さ・信憑性 (naturalness and believability)、信頼感 (trust)、知覚タスクの困難性 (perceived task difficulty)、好ましさ (likeability)



おわり