

知的情報処理論（第8回）

2023年6月6日（火）
産業科学研究所
駒谷 和範

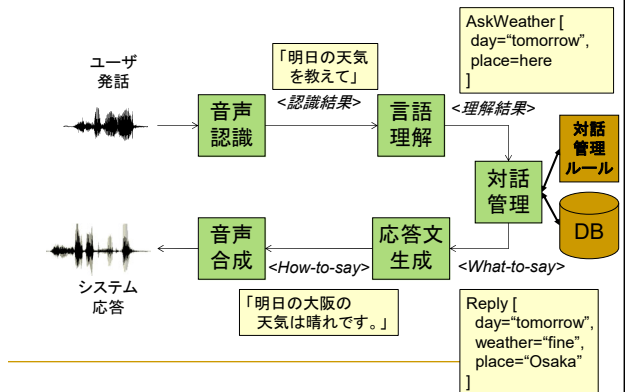
レポート（第2回）は来週締切です

- 締切：6月12日（月）
 - 考察や説明を、他人が読んでわかる日本語（または英語）で書くこと。
 - 実行結果のみを送りつけてきた場合は、極めて低く評価する、または受理しない。
- レポートの目的：
 - 講義の内容を、自分で手を動かして確認し、考察としてまとめることにより、自らの理解を深める（それを文書で伝える）
 - 「自主的、多面的に検討を行ったものには加点」
 - オリジナリティ歓迎、図示（可視化）歓迎
 - いろいろ試せる課題（第2回なら問2）は加点のチャンス

本日（第8回）の内容

- 第1回 人工知能の歴史
- 第2回第3回第4回 機械学習の基礎
- 第5回 対話システムの構成と分類、言語理解
- 第6回 （大規模）言語モデル
- 第7回 言語理解の利用法、対話管理
ニューラル対話システム
- 第8回 言語内容以外に必要な要素
人工知能と社会

音声対話システムの基本構成（対話型）



言語内容以外に必要な要素 (1) 発話タイミング

現状の音声対話インタフェースにおける暗黙の仮定

1. 発話は必ず交互に行われる
 2. ユーザ発話の最中にポーズはない
 3. 複数の指示を同時に（一発話中で）言わない
- 人間の場合
 - 特に親しい人との音声対話では、**発話はかなりオーバーラップ**
 - 「聞いている」というシグナルを示す相づち（「ええ」）
 - （遅延のある）オンライン対話における違和感
 - 音声特有：テキストチャットではターン構造は明らか

「対話＝トランシーバ」ではない

- Push-to-talk
押してからしゃべる
＝音声を入力するタイミングを明示
- 対話は、必ずしも二者が交互に話すとは限らない
 - あいづち、発話への割り込み、etc.
- ロボットにはボタンはない



<http://www.apple.com/jp/ios/siri/> より

対話における「発話」の定義

1. 物理的な定義
 - 例:「400ミリ秒以上の無音区間で区切られた区間」
 - 無音区間: 音響信号の振幅が一定値以下である区間
 - ただし促音「っ」は物理的には無音区間
2. 対話行為による定義
 - 意味的なまとまりを「一発話」とみなす
 - 「帰り・・・は新大阪までお願いします」
3. 話者交替による定義
 - 相手の発話に区切られた区間を一発話とみなす
 - 「どこまでですか?」→「新大阪、です」→「わかりました」

電子情報通信学会誌 (小特集: 音声
言語理解のこれまでとこれから)
Vol.101, No.9, pp.908-913, 2018.

小特集 7.

円滑な対話進行のための音声からの情報抽出

Obtaining Information from Spoken Utterances for Smooth Dialogues

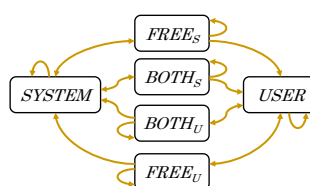
駒谷和範

Abstract

ユーザの音声発話には、その意味内容以外にも様々な情報が含まれる。とりわけ音声対話システムでは、ユーザとシステム間の発話のタイミングの関係や、更には発話が誰に向けられたものかによって、それを受け取ったシステムが取るべき挙動は異なる。対話を行うには、ユーザ発話に対してシステムが応答を始めるまでに大きな遅延が生じたり、ユーザとシステムの発話が意図せず重なったりするのを避ける必要がある。更には、システムが話すべきでないときに応答を始めるのも望ましくない。本稿では、このような円滑な対話の実現のために必要な技術について概説する。

キーワード: 音声対話システム, 話者交替, ターンテイキング, 発話区間検出, 受話者検定

ターンテイキングの状態モデル [Raux 2016]



- SystemとUserのどちらかがフロア(発話権)を持っているか
- 添字(S, U)は前状態

例:

- 無音区間を伴う話者交替

$$SYSTEM \xrightarrow{(R, W)} FREE_S \xrightarrow{(W, G)} USER$$
- 割り込みの失敗

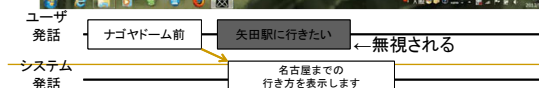
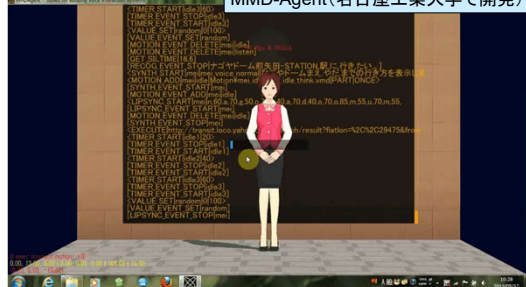
$$USER \xrightarrow{(G, K)} BOTH_U \xrightarrow{(R, K)} USER$$

$$SYSTEM \xrightarrow{(K, G)} BOTH_S \xrightarrow{(K, R)} SYSTEM$$

- ()内は、システムとユーザの行動
- G (grab) フロアを取る
 - R (release) フロアを手放す
 - W (wait) フロアを取らずに待つ
 - K (keep) フロアを維持する

発話の途中でシステムが話し始める例[堀田+13]

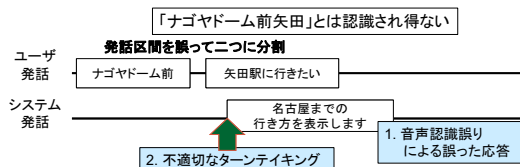
MMD-Agent(名古屋工業大学で開発)を使用



11

発話区間の修復

発話区間検出の誤りが、音声認識誤りの原因となる場合がある



⇒ 音声認識結果とターンテイキングの両方を修復

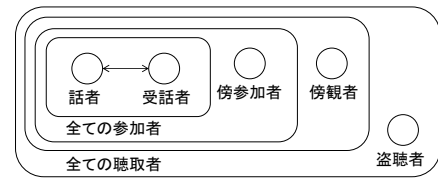
1. 発話区間を結合して再度音声認識を実行
2. ターンテイキングの管理機構の拡張, フィラーの生成

12

対話でロボットが理解すべきこと

- 参加者の同定
 - エンゲージメント(Engagement)の推定:
 - どの程度その人がその対話に参加しているか
 - 存在(位置)
 - 体の向き、顔の向き(視線方向)
 - うなずき、発話
- 受話者推定(誰に向けられた発話か)
 - 話者の顔の向き(視線方向)
 - 発話内容

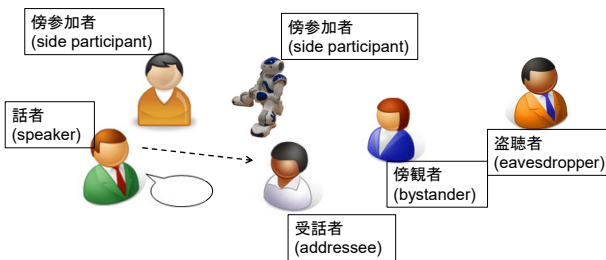
多人数対話における参加構造 [H. Clark, 1996]



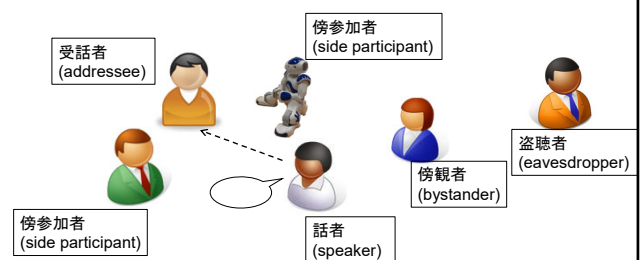
Herbert H. Clark, Using Language,
Cambridge University Press, 1996
<https://doi.org/10.1017/CBO9780511620539>



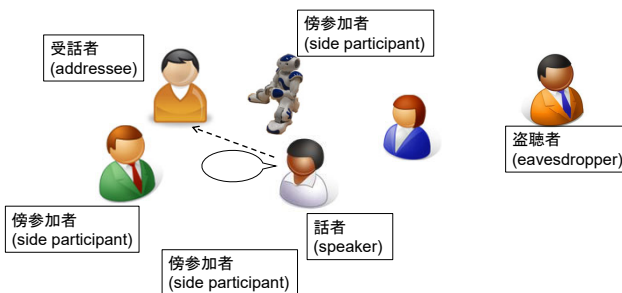
多人数対話の参加構造



多人数対話の参加構造



多人数対話の参加構造



多人数対話システムの難しさ

- 対話管理
 - 一対一の場合: 「聞かれたら答える」でよい
 - 音声が入力されたら、その結果に基づいて何か出力する
 - 多人数の場合: **役割の認識が必要**
 - 誰に向かって話すべきか
 - いつ話すべきか
 - (何を話すべきか) ← これは一対一の場合も同様
 - 状態追跡?

言語内容以外に必要な要素 (3) 感情

感情 (affect)

- 一時的な情動 (emotion)
- 長時間持続する状態
 - 気分 (mood)
 - 性格特性 (personality trait)
 - Big Five: 外向性, 情緒安定性, 協調性, 誠実性, 開放性
- 1997年 Picard “Affective Computing”

Paul Ekmanのモデル

- 心理学者
- 基本6感情 [Ekman 92]
 - 喜び, 嫌悪, 驚き, 悲しみ, 怒り, 恐れ
- Facial Action Coding System (FACS)
 - Action Unit (AU)



AU17
顎先を上げる



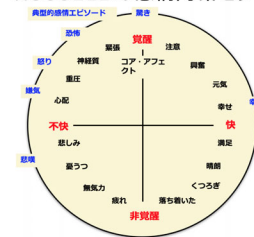
AU25
顎を下げずに唇を開く



AU26
顎を下げて唇を開く

Russellの感情円環モデル

RUSSELL の感情円環モデル



2軸

- Valence (快 - 不快)
- Arousal (覚醒 - 不覚醒)

Russell, J. A., & Barrett, L. Feldman (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, 74(5), 805-819.

- <https://note.com/fair/n/n8a0f600b2491>

内的状態とその表出

- 「感情が読み取れる」は本当？
- 機械が読み取っているのは「表出された感情」

https://doi.org/10.11517/jjsai.36.1_4

4 人工知能 第36巻 1号 (2021年1月)

【第2巻】Affective Computing

Affective Computingの研究分野：学際的視点

The Field of Affective Computing: An Interdisciplinary Perspective

Gratch, Jonathan 南カリフォルニア大学リサーチ・テクノロジー・イノベーション・研究所
Institute for Creative Technologies, University of Southern California
jonathan@ict.usc.edu

翻訳：アカデミアジャパン株式会社
AcademiaJapan

監訳：中田 裕樹 (横浜国立大学)
Kazunori Nishimoto (Yokohama National University)

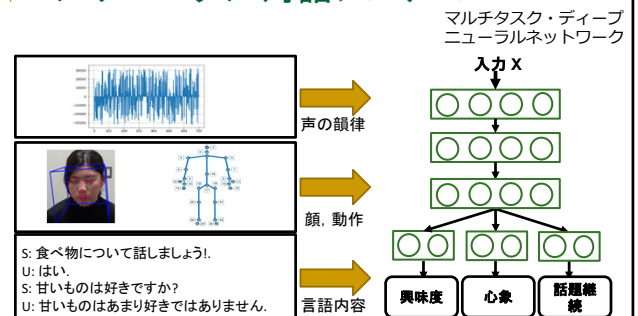
Keywords affective computing, mind, society, bias, ethics.

人工知能, Vol. 36,
No. 1 (2021年1月)

1. はじめに

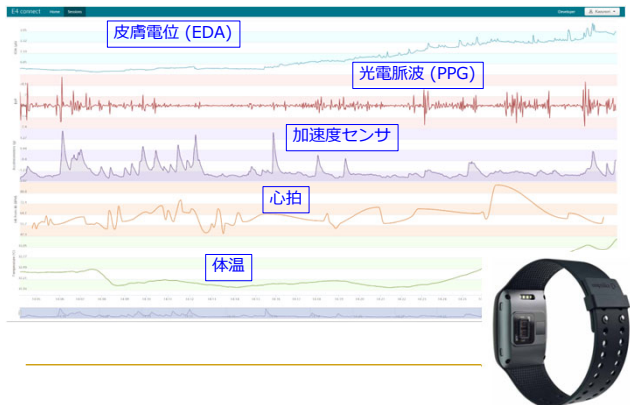
本号の一つである「本誌の創刊 2000年12月12日現在」
また、「Affective Computing」に関する論文を掲載する。

マルチモーダル対話システム



Yuki Hirano, Shogo Okada, Haruto Nishimoto, Kazunori Komatani:
Multitask Prediction of Exchange-level Annotations for Multimodal Dialogue Systems.
International Conference on Multimodal Interaction (ICMI), pp.85-94, 2019.
(Best Paper Runner-ups Award)

生体センサ (Empatica社 E4)



生体センサを使った実験

- 本人の心象(振り返りアノテーション結果)を生体信号を使って予測
 - 第三者アノテーションの平均で予測: 63.0%
 - 皮膚電位を使った予測によりこれと同等の性能

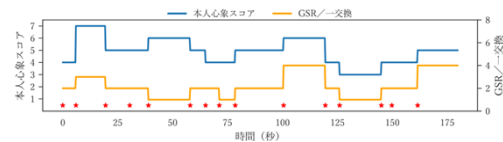


図5: 実験参加者本人が付与した心象とGSR特微量の時間変化の例

Shun Katada, Shogo Okada, Kazunori Komatani:
Effects of Physiological Signals in Different Types of Multimodal Sentiment Estimation.
IEEE Transactions on Affective Computing, (accepted), 2022.
<https://doi.org/10.1109/TAFFC.2022.3155604>

言語内容以外に必要なこと (4) 語用論 (PRAGMATICS)

発話行為理論 (Speech Act Theory)

- J.L.Austin (1962)
- 発話を行為として捉える: 3種類の階層
 1. 発語行為 (locutionary Act)
ある言語の語彙・文法を使って発話する行為
 2. 発語内行為 (illocutionary Act)
発話により, 主張, 命令, 依頼, 警告, 約束などを行う行為
 - この行為の効力を illocutionary force と呼ぶ
 3. 発語媒介行為 (perlocutionary Act)
発話の結果として聞き手に及ぼす行為

発話行為の階層の例

閉店間際のバーで「もうラストオーダーですよ」

- 発語行為:
上記の文を発声するという行為
 - /mo:rasuto.../ という音声を発する行為
- 発語内行為:
ラストオーダーの時間であるという陳述
- 発語媒介行為:
最後にもう一杯注文するのを促す行為

間接発話行為 (indirect speech act)

- 食事の場面で「それは塩ですか？」
 - 表層的には Yes/No 質問
 - 語用論的には「塩を取ってください」という依頼

Griceの会話の公準 (maxim) (1967)

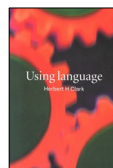
- 協調性の原理を満たすための原則
 - ① 質の公準 (maxim of quality)
真なる発言をする
 - ② 量の公準 (maxim of quantity)
過不足のない情報を伝える
 - ③ 関係の公準 (maxim of relation)
関連した事柄を話す
 - ④ 様態の公準 (maxim of manner)
明確に, 簡潔に, 順序立てて話す
- 含意 (implicature): 明示されていないけれども伝わる

会話の公準を(わざと)逸脱した例

- 自分の(本当の)子どもを指しながら
「僕は彼の法律上の父です」
 - 不自然
 - これは量の公準を満たしていないため、不自然な「含み」が生じているか

言語内容以外の様々な情報

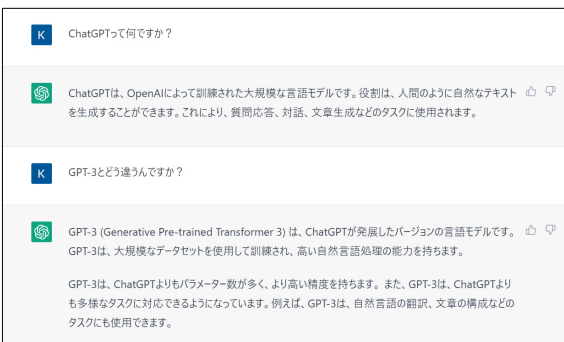
- 対話の理論
 - 「対話システム」(2章)
 - “Using Language”
H. H. Clark (1996)
- 非言語情報・パラ言語情報
 - 「音声は何を伝えているか」



人工知能と社会

ChatGPT (2022年12月)

<https://chat.openai.com/chat>



画像生成技術 (StyleGAN) (2019)

- <https://www.gwern.net/Faces>

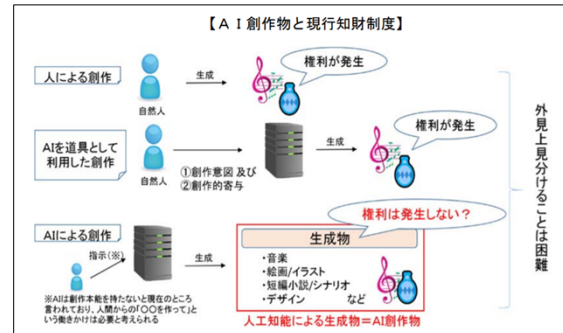


Stable Diffusion (2022)

- Stability AI (ドイツのスタートアップ)
 - “A pikachu fine dining with a view to the Eiffel Tower”
<https://huggingface.co/spaces/stabilityai/stable-diffusion>



「人工知能が書いた」作品の著作権は？



「次世代知財システム検討委員会報告書～デジタル・ネットワーク化に対応する次世代知財システム構築に向けて～」より引用
https://www.kantei.go.jp/jp/singi/tuteki2/tyousakai/kensho_hvoka_kikaku/2016/jisedai_tizai/hokokusho.pdf

著作権問題

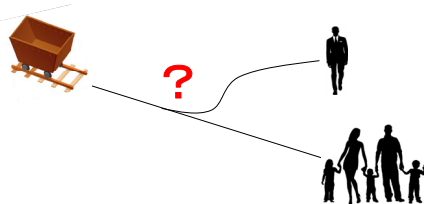
- 生成されたものの著作権は？
- Webにあるデータは学習データとして勝手に使っていい？
 - 日本では平成30年著作権法改正でOK
 - 「著作物に表現された思想又は感情の享受を目的としない利用」を認める
 - 当時、認識系の機械学習に使われることが念頭に置かれていた
 - 生成型AIの勃興により議論が再燃？
 - その利用が著作権者の利益を不当に害する場合はダメ

人工知能の製造者責任

- 製造物責任法(PL法)
 - 製造物の欠陥により被害が生じた場合に、製造者等の損害賠償の責任について定めた法律
 - メーカーが作った人工知能(例えば自動運転車)が事故を起こした場合、誰の責任？

トロッコ問題のパラドックス

- 倫理学における思考実験
- ある人を助けるために、他の人を犠牲にするのは許されるか



5人が1人か、どちらを救う？ 自動運転車が直面する「トロッコ問題」【けいざい百景】

7/23(土) 10:30 配信 1402

JIJI.COM 時事通信社

2022/7/23(土)

<https://news.yahoo.co.jp/articles/519bd2b3b156d88aff1b6680563874ffb68fb651>



「自動運転における「トロッコ問題」の例

Mini Quiz #2

- 完全自動運転車の事故は誰のせい(≡事故の賠償責任は誰が負うべき)?
その理由は?
 - ユーザ(車の所有者, 乗っていた人)?
 - その人工知能(その車一台)?
 - 車や自動運転プログラムを作ったメーカー?
 - 学習データを集めた人?
 - 国?
 - 誰のせいでもない?
- そもそもそんなものを開発してはいけない?

人工知能と兵器

- 兵器は「リスクがある」ではなく「既に禁止」
 - 自律型致死兵器システム (LAWS)
Lethal Autonomous Weapons Systems
 - 兵器は「人間が判断の決定に関わること」が前提
 - 有意の人間の判断 (meaningful human control) がシステムに期待できない場合、攻撃を実行する主体の責任が不明瞭
- 一方、自律的な防衛システムは既に存在
 - イスラエルのミサイル防衛システム アイアンドーム (Iron Dome)
- 「自律的」の定義は?
 - 自動運転車で歩行者を検知する技術と、攻撃対象の人を検知する技術は同じ

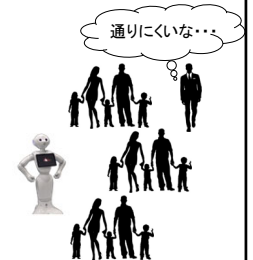
「法的にOK=やっている」ではない

- 2017年5月 人工知能学会 全国大会論文炎上事件
 - pixivで公開されていた創作物(二次利用小説; R-18指定)を、わいせつ表現検出の対象データとして利用
 - 出典(当該サイトURLや作者のペンネーム)を論文に明記
 - 当該コミュニティでは「晒し上げだ」として炎上
 - 学術論文における引用の作法としては適切
 - pixivの当該コミュニティへの影響は大きい

学会コミュニティではOK(むしろ義務)である作法が、別のコミュニティでは受け入れられないことがある
- 2019年8月 リクナビ社による内定辞退予測の販売

社会に受け入れられる?

- 公共の場での、ロボットを使った公開実験
 - 実験参加者は楽しんでいる
 - 実験に参加しない人には邪魔
社会ではいろんな利害関係者が存在
- 法的には問題ない
 - 社会に受け入れられるか?
 - 倫理の問題/説明責任
 - 新しい技術(人工知能)を設計する側が、社会に与える影響を考慮することが求められ始めている
⇒ ELSI (Ethical, Legal and Social Issues)



「人工知能と社会」のまとめ

- 人工知能が社会制度に与える影響
 - 人工知能の著作権
 - 新たな権利
 - 他人(クリエイター)の権利を侵害する恐れ
 - 製造者責任
 - リスクを考えた設計
 - 兵器
 - デュアルユースへの懸念、自律的かつ致死的なものは既に禁止
- 開発者のモラル・社会との対話
 - お互いに信用が必要
 - 「自分が想像できないことがある」という想像力
 - 「グレーなことは一切するな」も委縮につながる
 - 説明責任