第9回対話システムシンポジウム　チュートリアル

マルチモーダル対話システムにおける社会的信号処理とエージェント技術

成蹊大学理工学部情報科学科
中野有紀子

本日の内容

- マルチモーダル対話システムとは？
- 社会的信号処理
  - 会話参加者の特性
  - 感情
- マルチモーダル情報に基づく対話制御
  - フロアマネジメント
  - 参加態度の推定
- エージェント技術
  - ジェスチャ生成
  - 表情生成
- 統合システム
- まとめ

マルチモーダルとは

- モダリティとは
  - 情報のタイプ，あるいは情報の表現形式
  - 感覚のモダリティ：感覚の形式（視覚，触覚等）
  - コミュニケーションのチャンネル
- モダリティの種類
  - 自然言語（音声言語，文章） → 言語
  - 視覚情報（画像，映像）
  - 聴覚情報（音声，音，音楽）
  - 触覚
  - 匂い，味
  - 生理指標（心拍，発汗）
  - その他（脳波，fMRI）

非言語

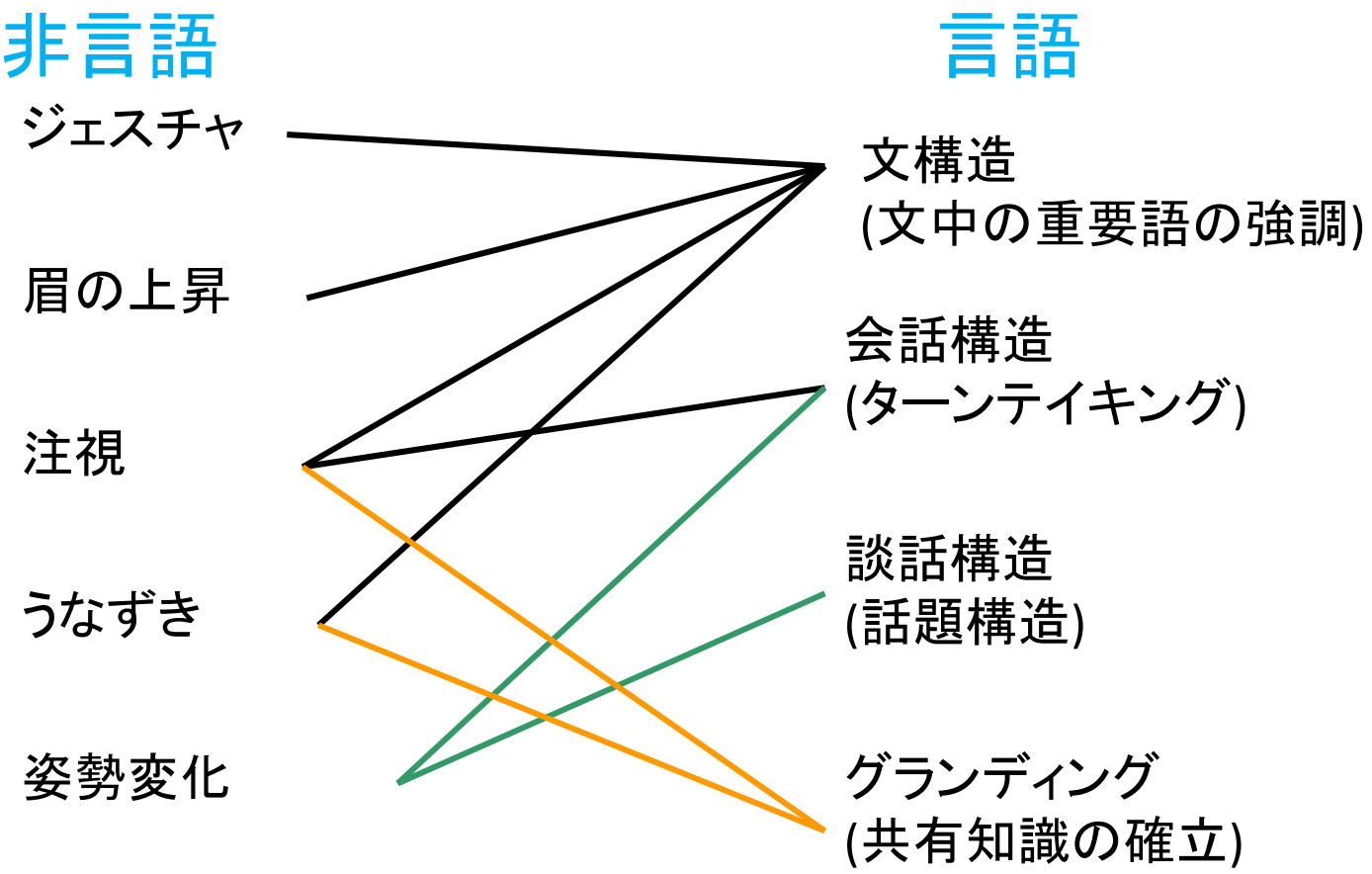
【参考】
(マルチ)メディア
情報伝達・保存の手段・システム

言語と非言語コミュニケーションシグナル

- 言語情報
  - テキストや音声により表現される言語情報
- 非言語情報
  - 身体的表現(ジェスチャ，視線，姿勢，表情など)
  - 韻律情報(声の高さ，大きさ，抑揚など)
- 機能
  - 言語情報：命題的機能
    - 命題内容を表現
  - 非言語情報：インタラクション機能
    - 言語情報に付随して，言語情報の伝達を補助する

ジェスチャ

言語と非言語コミュニケーションシグナルの関係

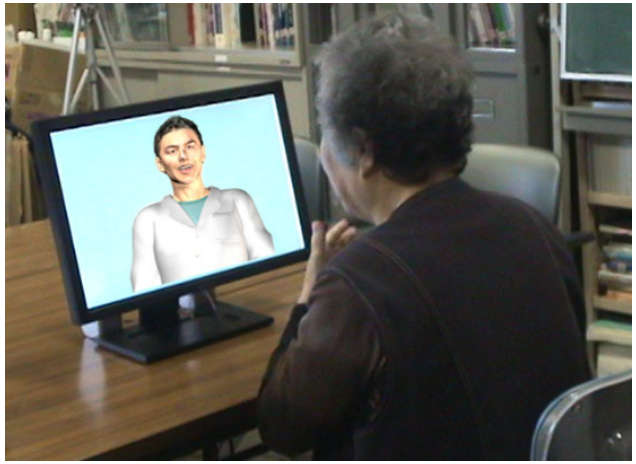


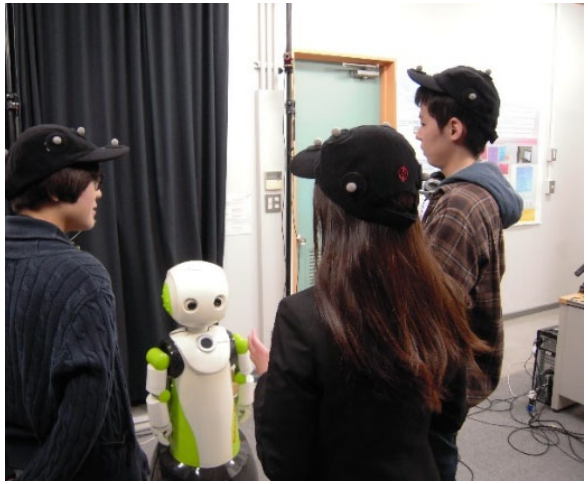
感情

社会的関係

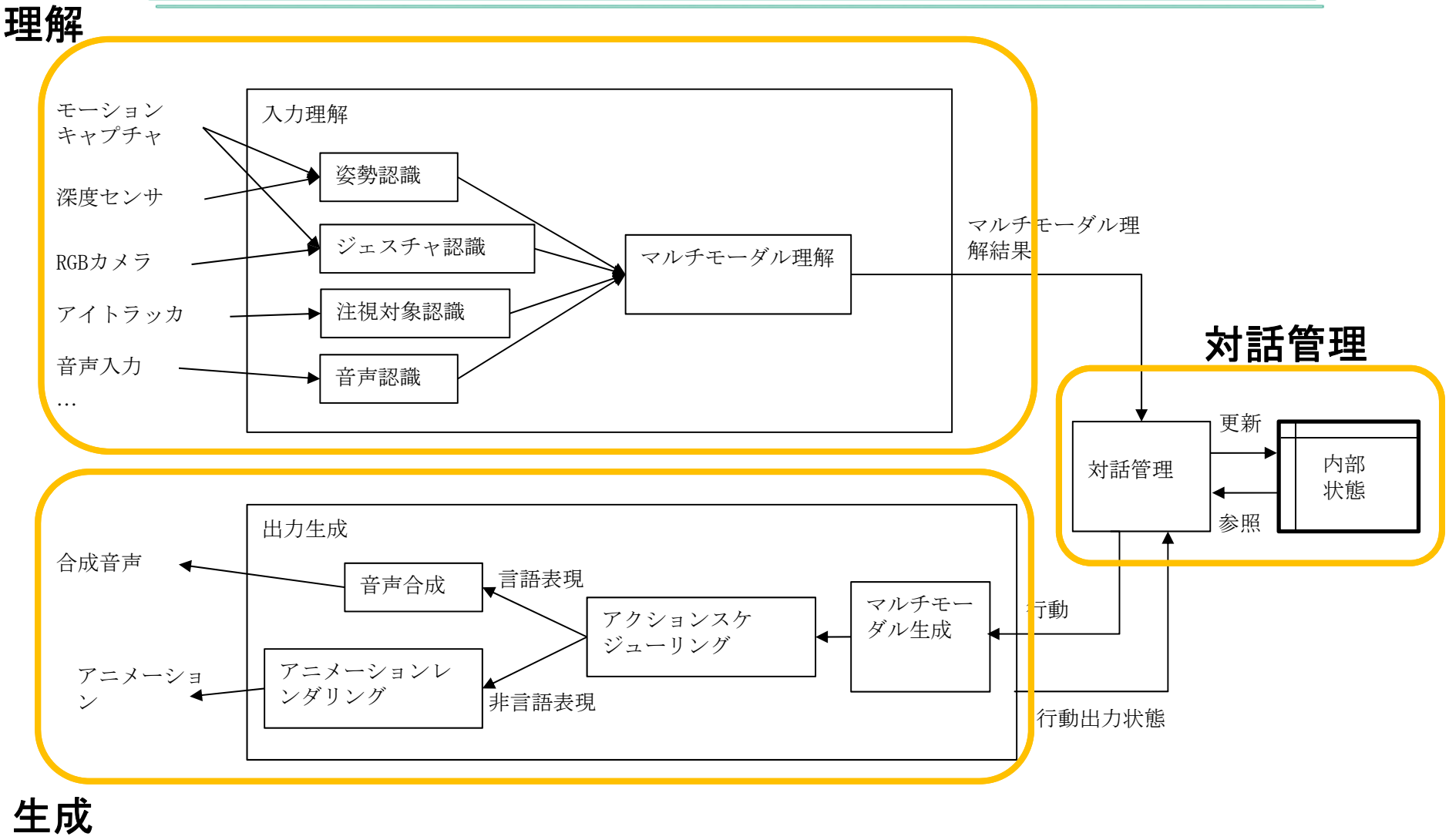
マルチモーダル(ヒューマノイド型)対話システム

- 対面会話を人と人工物の間で実現することを目指す
  - 2者間対話
  - 多人数(マルチパーティ)対話
- ヒューマノイドである意味
  - 身体性の意味
    - 直観性，頑健性，自然性において優れている
  - 身体性の効果
    - 人工物であるほうが話しやすい(自己開示)等

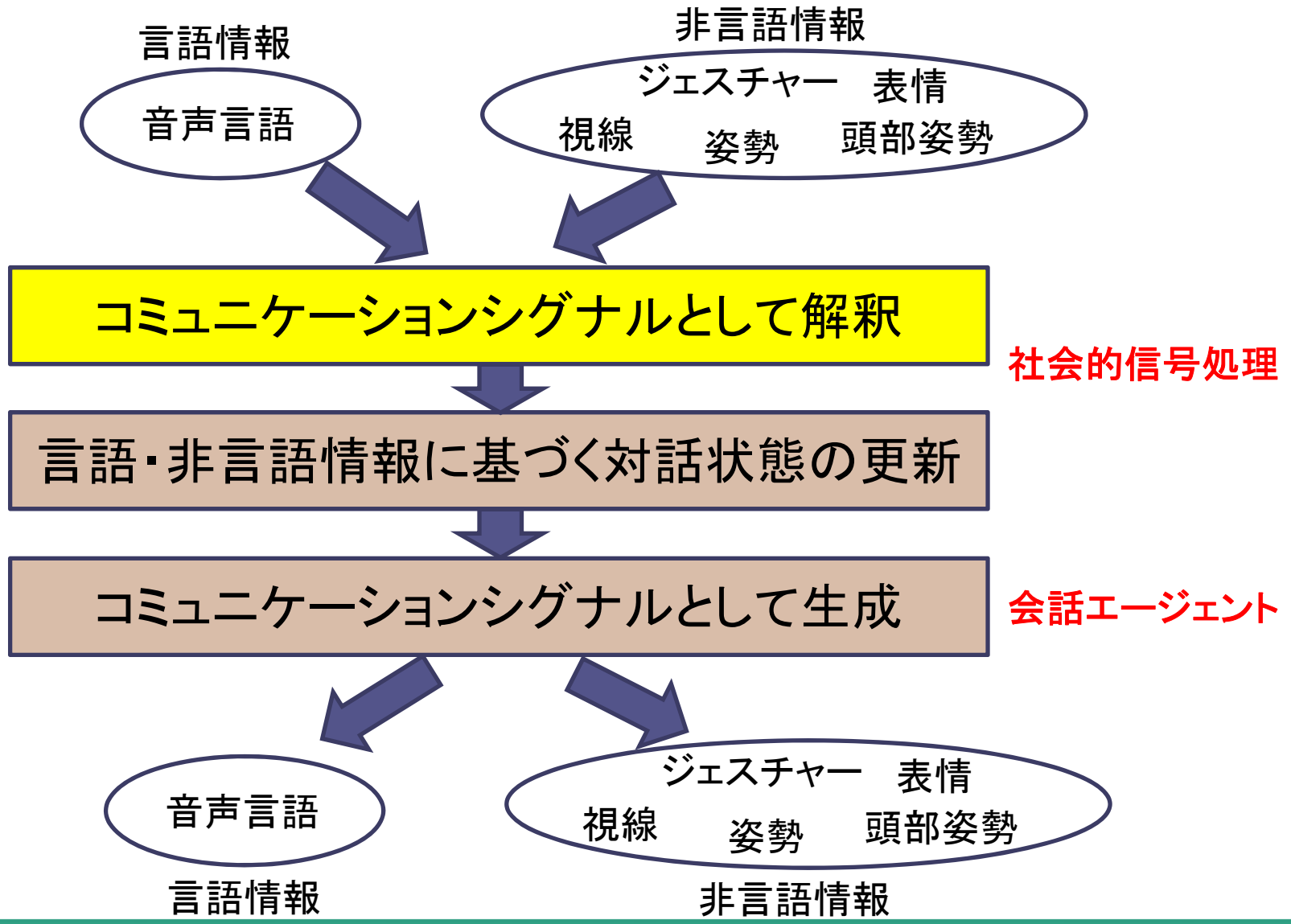




マルチモーダル対話システムアーキテクチャ



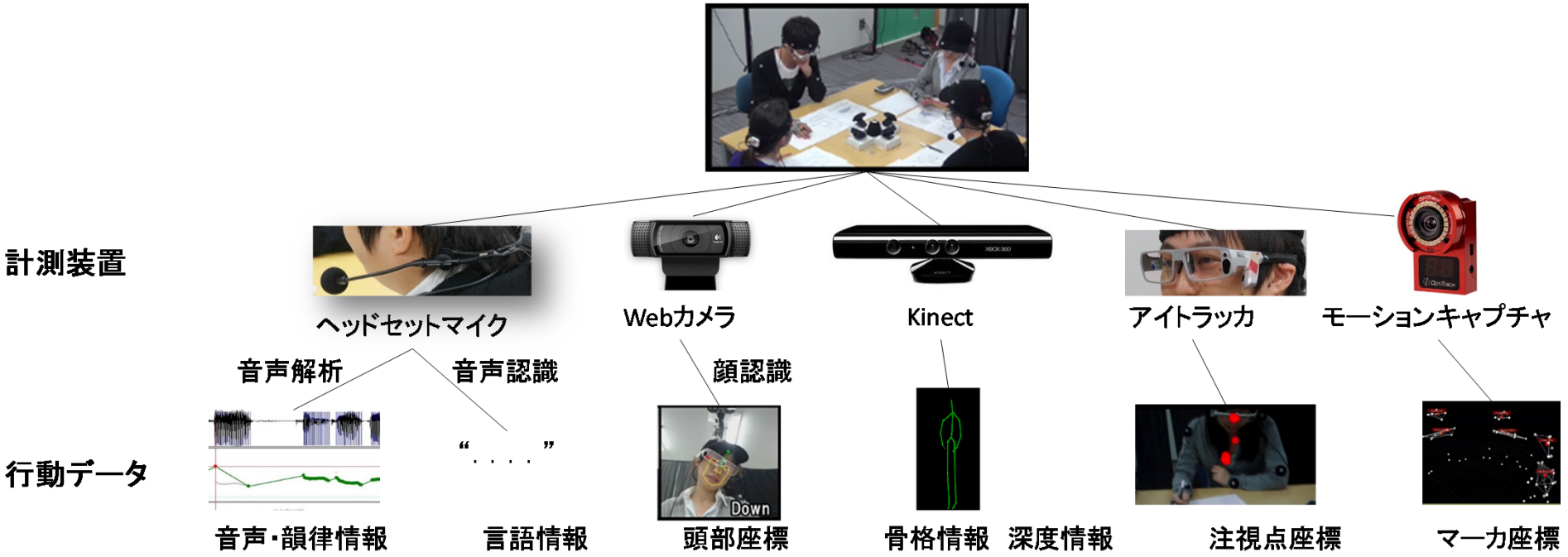
コミュニケーションシグナルの理解



社会的信号処理

- 正直シグナル(Honest signal) [Pentland, 2008]
  - 他者とのインタラクションにおいて，無意識に処理される／意識的にコントロールできない微細な行動パターン（社会的信号）
- 社会的信号処理とは：人の社会的信号をコンピュータにより認識・理解する研究・技術分野 [Vinciarelli et al. 2009]
  - 統計的手法や機械学習を適用する

行動データの収集：機材と計測データ



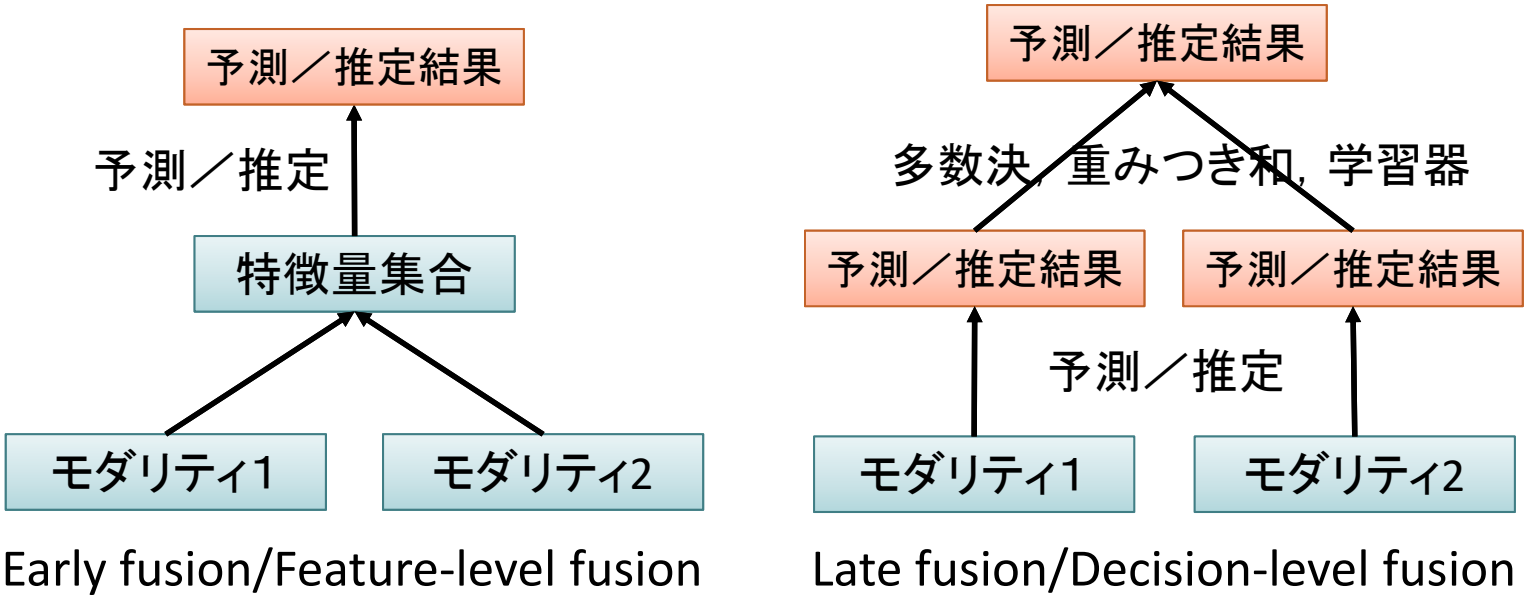
- 最近よく使われているソフトウェア
  - OpenSmile：音声解析
  - OpenFace：顔動作特徴量抽出(AU，頭部座標)
  - OpenPose：骨格情報

言語・非言語特徴量

- マルチモーダルコーパス
  - AMI（Augmented Multi-party Interaction）[Carletta et al. 2005]
  - ELEA [Sanchez-Cortes et al. 2013]
- 非言語特徴量
  - 発話ターン：発話総量，発話ターン回数，平均発話ターン長
  - 韻律：ピッチ最小値／最大値／中央値／標準偏差，インテンシティ最小値／最大値／中央値／標準偏差
  - 頭部動作：頭部動作総量，頭部動作回数，頭部動作平均継続長，X軸方向の頭部動作標準偏差，y軸方向の頭部動作標準偏差
  - 胴部動作：胴部動作総量，胴部動作回数，胴部動作平均継続長，胴部動作標準偏差
  - 注視：注視量，被注視量
- 言語特徴量
  - 単語数，品詞，発話中の位置，対話行為ラベル
- マルチモーダル特徴量：共起関係を特徴量とする
  - 発話×注視：発話中の注視量，傾聴中の注視量，，発話中の被注視量，傾聴中の被注視量，傾聴中の注視量に対する発話中の注視量の比率
  - その他対話参加者間の行動の共起関係など
  - 共起関係の組み合わせは爆発する！

マルチモーダルフュージョン(1)

- 単一モダリティの情報のみによる推定よりも，複数モダリティの情報を用いるほうが推定性能がよい



Early fusion/Feature-level fusion　　Late fusion/Decision-level fusion

Early fusionとLate fusionのハイブリッドもある

性格印象推定

- 性格印象推定：性格特性に関する他者の印象を推定
  [Aran et al. 2013][Okada et al. 2015, 2018] Fang et al. 2016]
- Big Five 性格特性(OCEAN)
  - **O**penness to Experience（経験への開放性）
  - **C**onscentiousness（勤勉性）
  - **E**xtraversion（外向性）
  - **A**greeableness（調和性）
  - **N**euroticism（情緒不安定性）
- 特徴量選定
  - 最初はできるだけ多くの特徴量を設定し
  - 相関分析やt検定等で単独要因での予測力・有用性を検証
  - 有用な特徴量のみを用いてモデル学習
- 性格印象の推定：
  - Ridge回帰による性格印象スコアの推定
  - 2クラス(高／低)分類

ELEAコーパス

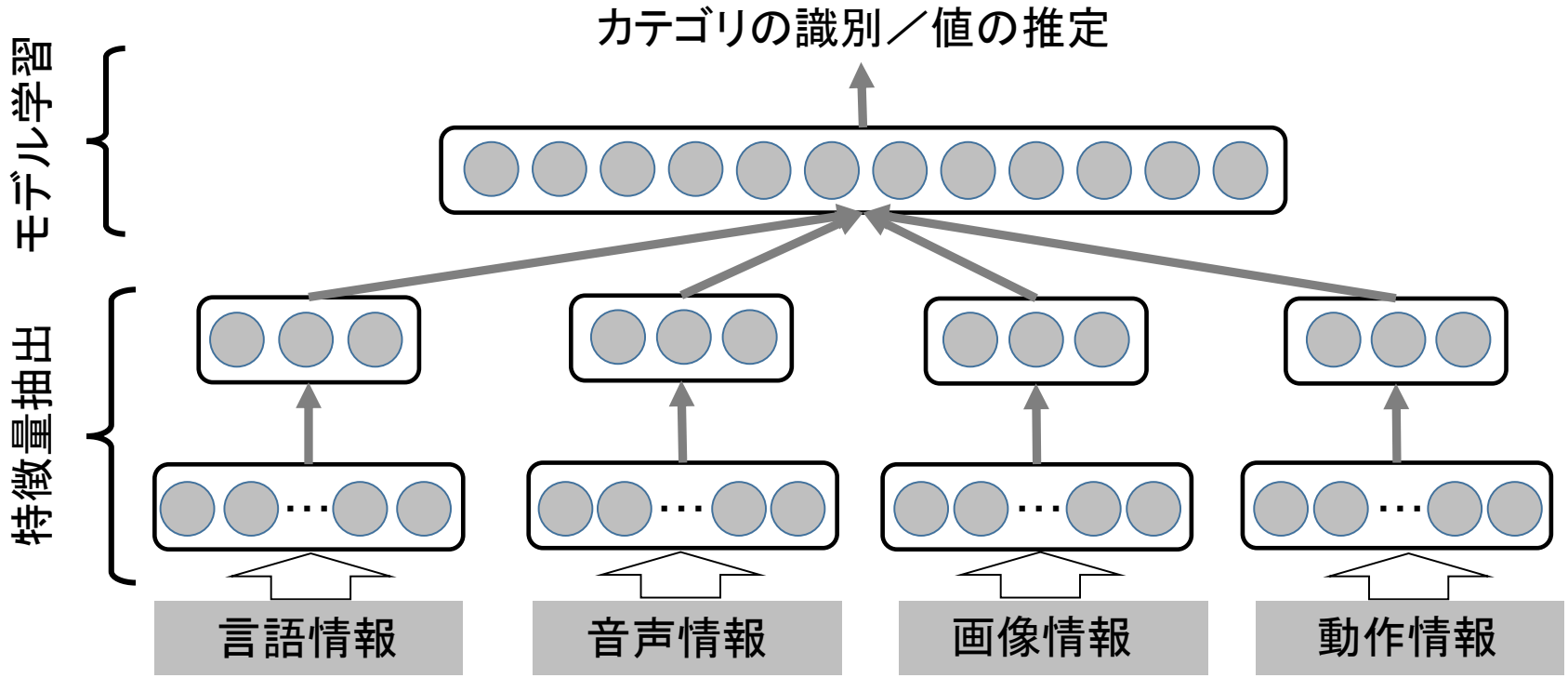
[Aran et al. 2013， Figure 2]

2クラス分類の性能(%)

	O	E	A
[Aran et al., 2013]	47.1	74.5	52
[Fang et al., 2018]	61.76	65.54	62.75

Aran, O. and Gatica-Perez, D. (2013). One of a Kind: Inferring Personality Impressions in Meetings. the 15th ACM on International conference on multimodal interaction (ICMI2013): 11-18.

マルチモーダルフュージョン (2)

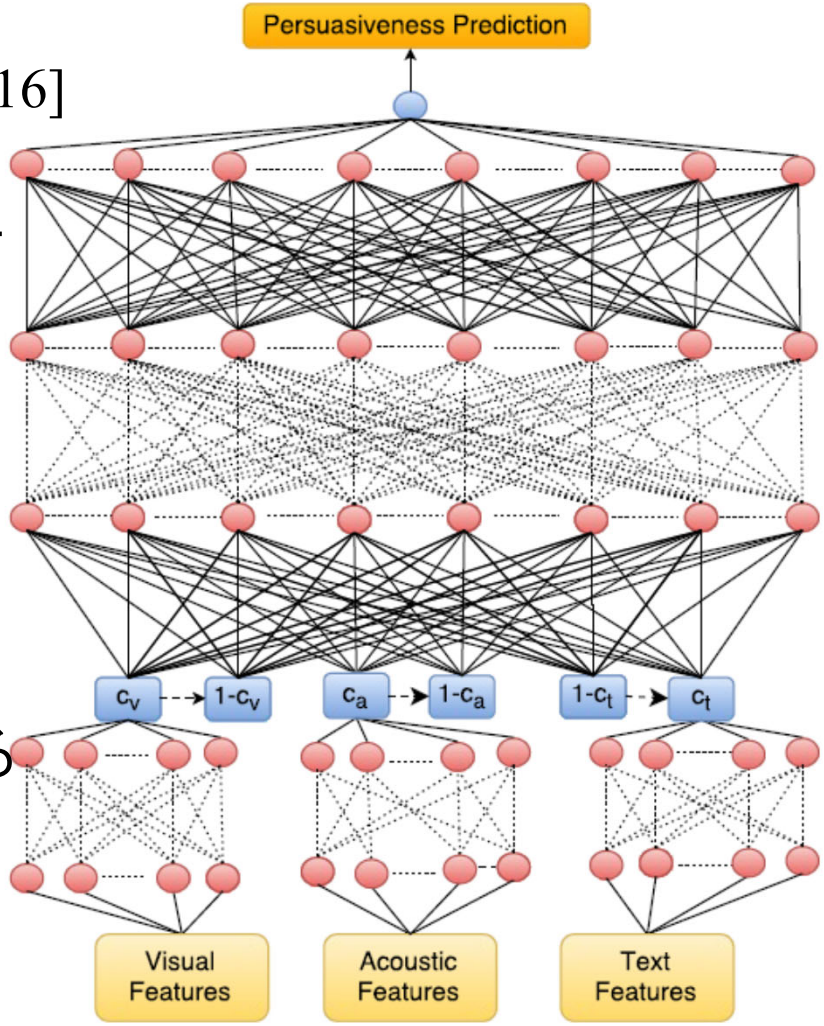


- モダリティによってデータの形式が異なる
  - テキスト：記号
  - 音声，画像：信号

深層学習では全てのモダリティをベクトルで表現
→マルチモーダルフュージョンが容易に！

説得力の推定

- 説得力の推定[Nojavanasghari et al. 2016]
- 学習データ
  - ソーシャルメディアでの映画レビュービデオ
- 特徴量
  - 顔表現特徴量(Visual)
  - 音声(Acoustic)
  - 言語(Text)
- 深層学習によるモデル学習
- Early fusion model：3種類のモダリティの特徴量を連結して入力ベクトルとする
- Late fusion model：
  - 各モダリティの深層学習モデルからの出力の平均値を予測値とする
  - 各モダリティからの出力を結合し，さらにこれを入力とした深層学習を行う

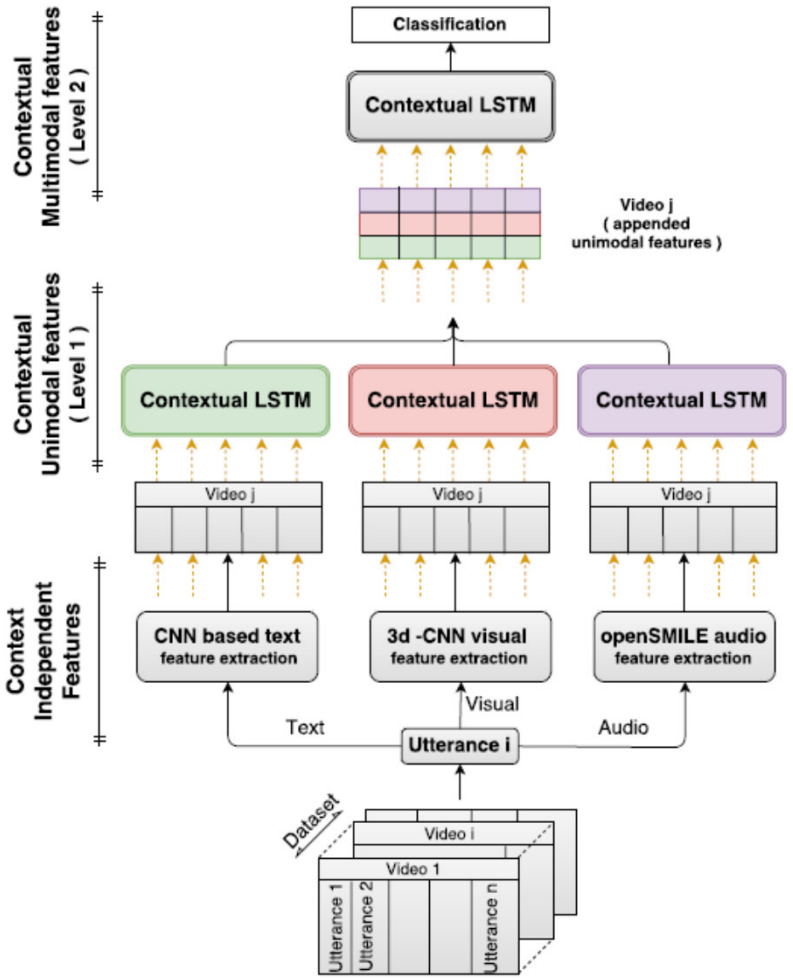


[Nojavanasghari et al., 2016, Figure 1]

Nojavanasghari, B., Gopinath, D., Koushik, J., Baltru, T., #353, aitis and Morency, L.-P. (2016). Deep multimodal fusion for persuasiveness prediction. Proceedings of the 18th ACM International Conference on Multimodal Interaction. Tokyo, Japan, ACM: 284-288.

感情認識

- Sentiment analysis
  - Positive/Negativeの極性を推定
- Emotion Recognition
  - 感情の種類(喜び，悲しみ，怒り，嫌悪等)を推定
- より複雑な深層学習の手法
  - 言語，韻律，画像
  - 特徴量選定なし
- Emotion Recognition in the Wild Challenge (EmotiW 2013-)
  - SEMAINEコーパス



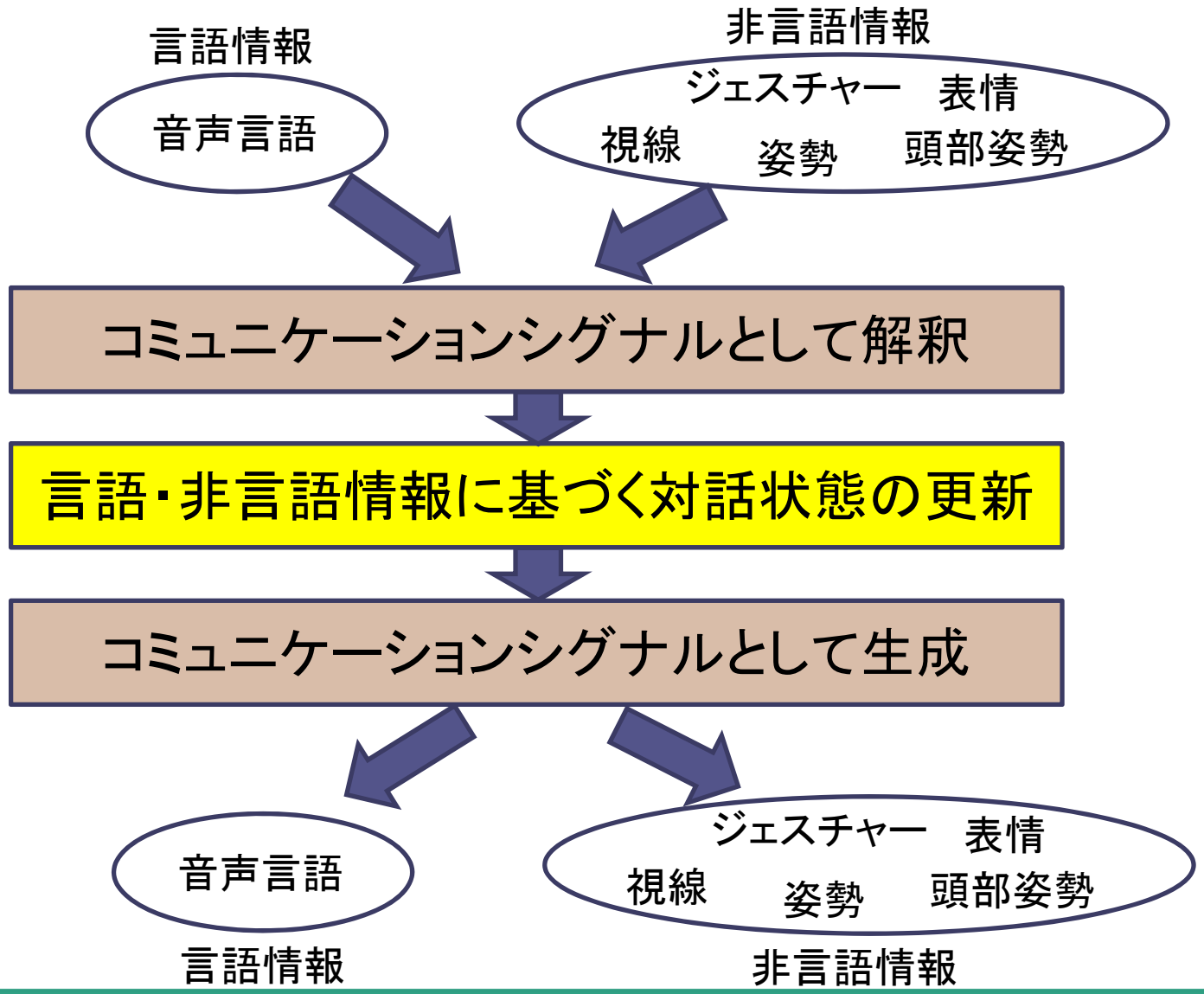
[Poria et al. 2017, Figure 2]

Poria, S., Cambria, E., Hazarika, D., Majumder, N., Zadeh, A. and Morency, L.-P. (2017). Context-Dependent Sentiment Analysis in User-Generated Videos. the 55th Annual Meeting of the Association for Computational Linguistics.

会話参加者特性推定の特徴量（まとめ）

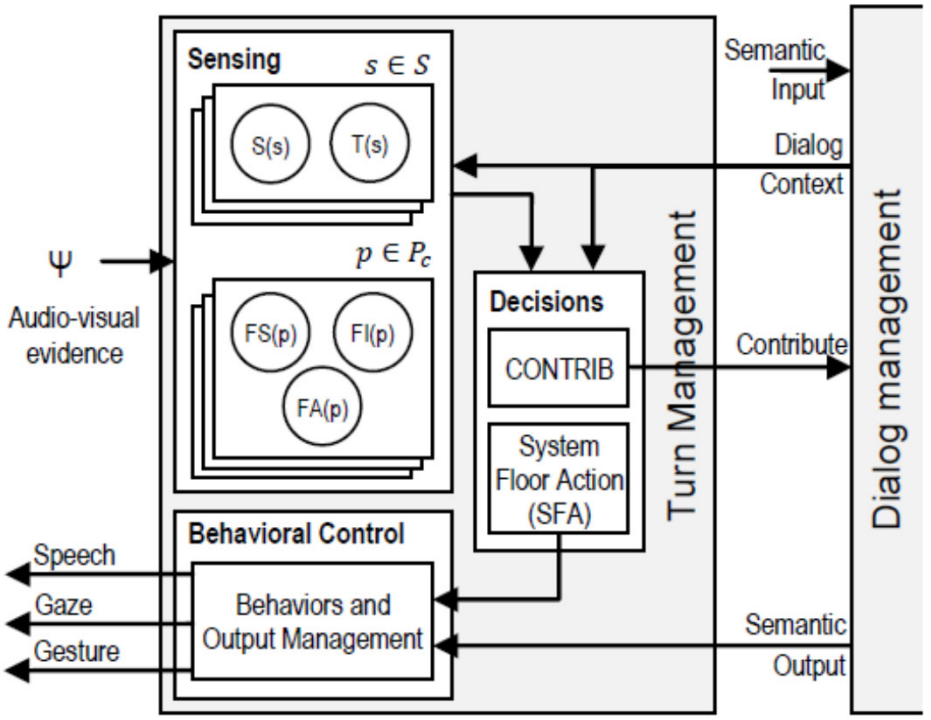
	会話特徴量	韻律	言語	頭部動作	ジェスチャ	胴部動作	注視行動	顔表情特徴量
優位性 [Sanchez-Cortes et al. 2013]	✔						✔	
性格印象 [Aran et al. 2013]	✔	✔		✔		✔	✔	
説得力 [Nojavanasghari et al. 2016]		✔	✔	✔				✔
感情 [Poria et al. 2017]		✔	✔					✔
共感性 [Kumano et al. 2012]							✔	✔
信頼性 [Lucas et al., 2016]	✔				✔		✔	✔
就職力 [Nguyen et al. 2013, 2014]	✔			✔	✔	✔		
コミュ力 [岡田 et al. 2016]	✔	✔	✔	✔				

マルチモーダル情報に基づく対話制御



マルチモーダル情報によるフロアマネジメント

- 信号取得(handcraftルール)
  - 各信号のソース
  - 各信号の送信対象者
  - 現在ターンを保持しているか否か (FS(p))
  - ターン取得意図／意欲があるか (FI(p))
  - どのフロアマネジメント行動を行っているか(FA(p))
- ターン交代決定(handcraftルール)
  - システムのコミュニケーション行動実行決定
  - フロアマネジメント行動の選択
- フロアマネジメント行動の決定
  - フロア保持：旧情報発話時は受話者に視線を向けない，新情報発話時はまんべんなく視線を向ける
  - フロア譲渡：ターン譲渡対象者に視線を向ける
  - フロア取得：現在のターン保持者とアイコンタクトを取り，それに成功したら話し始める



[Bohus et al. 2010, Figure 1]

Bohus, D. and Horvitz, E. (2010). Facilitating Multiparty Dialog with Gaze, Gesture, and Speech. The International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI'10).

ターン交代に関する予測

- マルチパーティ対話におけるターン交代予測の重要性
  - 2者対話と違い，次に誰が話すのかわからない
  - 視線など非言語情報によりターン譲渡シグナル，ターン交代シグナルの交換がなされる
- 注視行動のパターンに基づく，ターン交代の予測モデル[Ishii et al., 2016]
- 呼吸行動に基づくターン交代予測モデル[Ishii et al., 2014]
  - ターン交代の予測
  - 次話者予測
  - 次発話タイミングの予測
- LSTMを用いた音声と言語に基づくターン交代モデル [Roddy et al., 2018] [Skantze, 2017]
  - MapTaskコーパスの2者間対話

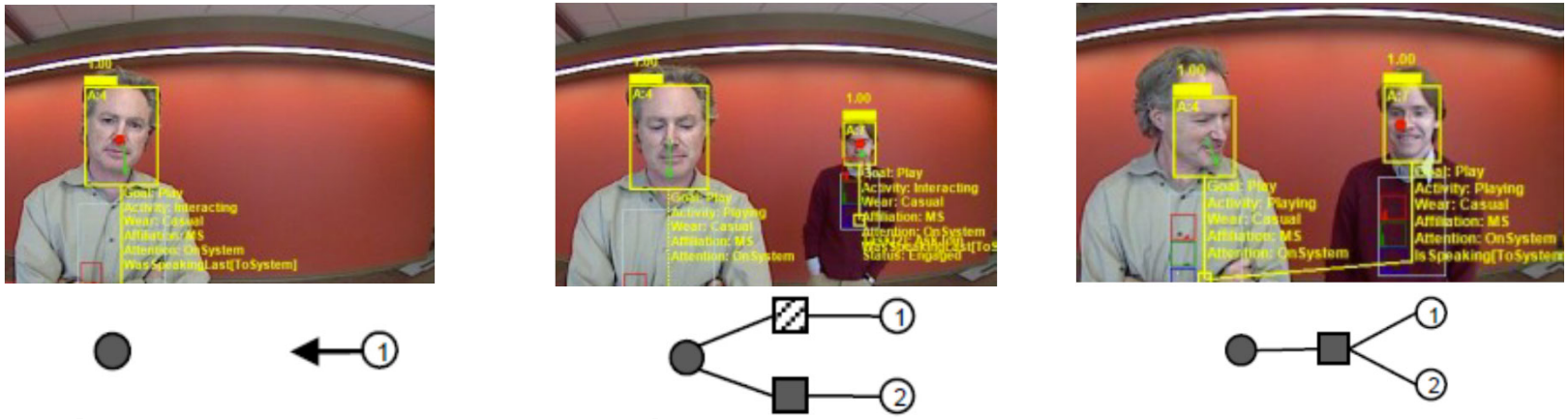
[Ishii et al., 2014, Figure 1]

Ishii, R., Otsuka, K., Kumano, S. and Yamato, J. (2014). Analysis of Respiration for Prediction of "Who Will Be Next Speaker and When?" in Multi-Party Meetings. the 16th International Conference on Multimodal Interaction (ICMI '14).

会話参加態度(engagement)の推定

- マルチパーティ対話システムにおけるengagementの問題
  - 複数人の会話参加行動や会話参加意図を認識
  - 複数人の参加状態の決定（誰が参加／不参加であるか？）
  - 参加状態によってシステムの振る舞いを決定（ジェスチャ，挨拶など）
- 会話参加状態管理機能を有するロボット[Bohus et al. 2009]
  - 発話，立ち位置などを計測
  - 確率モデルによる参加態度の推定：参加状態，参加行動，参加意図

一人目の参加　　システムが参加呼びかけ　　ユーザ同士の会話

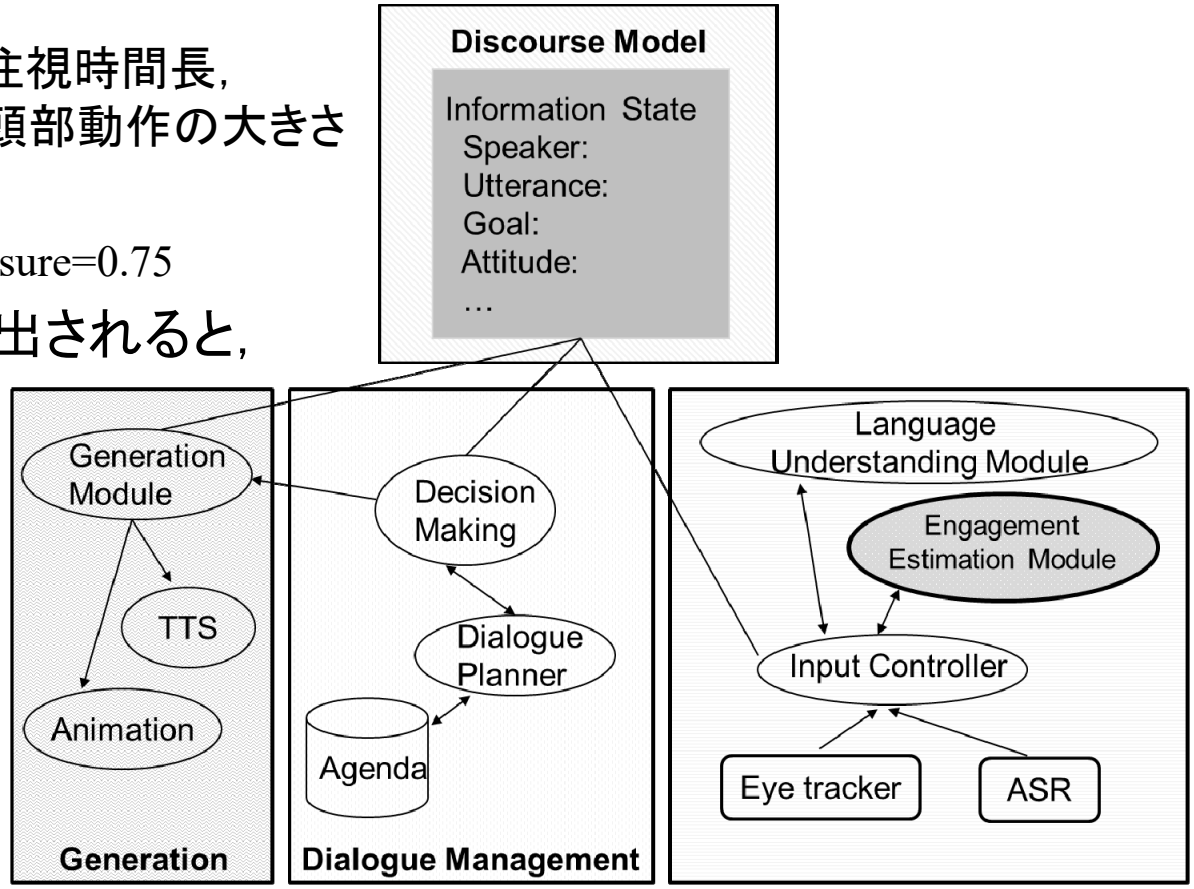


[Bohus et al. 2009, Appendix A]

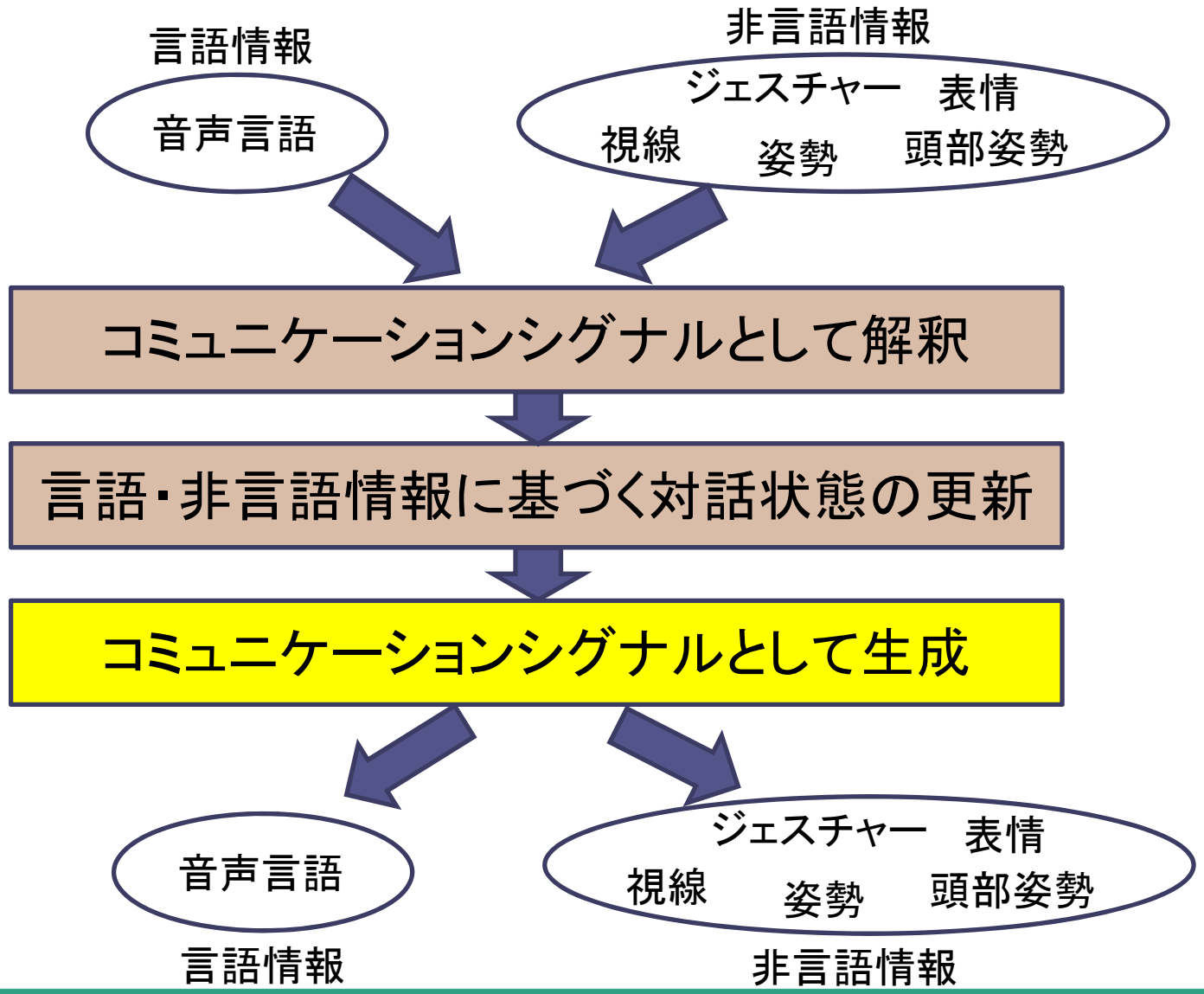
Bohus, D. and Horvitz, E. (2009). Learning to Predict Engagement with a Spoken Dialog System in Open-World Settings. SIGdial'09, London, UK.

参加態度の推定(Cont.)

- 注視行動に基づく会話参加態度の推定[Nakano et al. 2010]
- 積極的な会話参加態度であるか否かを推定
  - アイトラッカを使用
  - 注視対象遷移パターン，注視時間長，注視移動距離，瞳孔径，頭部動作の大きさ
  - 非積極状態の検出性能
    - SVMによる学習，F-measure=0.75
- 会話への関心低下が検出されると，エージェントが問いかけを行う



コミュニケーションシグナルの生成



発話に共起するジェスチャ

- ジェスチャ＝ハンドジェスチャ
- 音声言語との共起により意味が与えられる

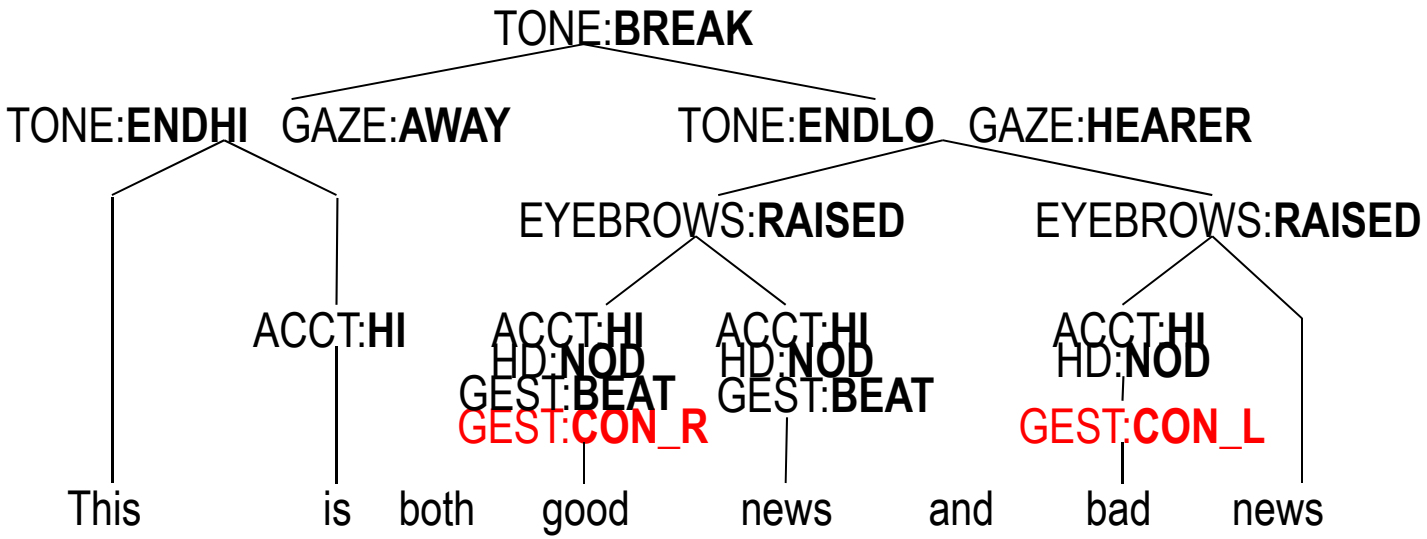
(1) 発話中の参照物と関連する
  - 発話中の指示詞と共起するポインティングジェスチャ

(2) 発話者の談話の視覚的な句読点として機能
  - 発話の韻律との関連性：韻律的に強調される語句にジェスチャが共起することが多い

(3) 対話の調整，構造化を支援
  - [Bavelas 1994]によるinteractive gestureの研究

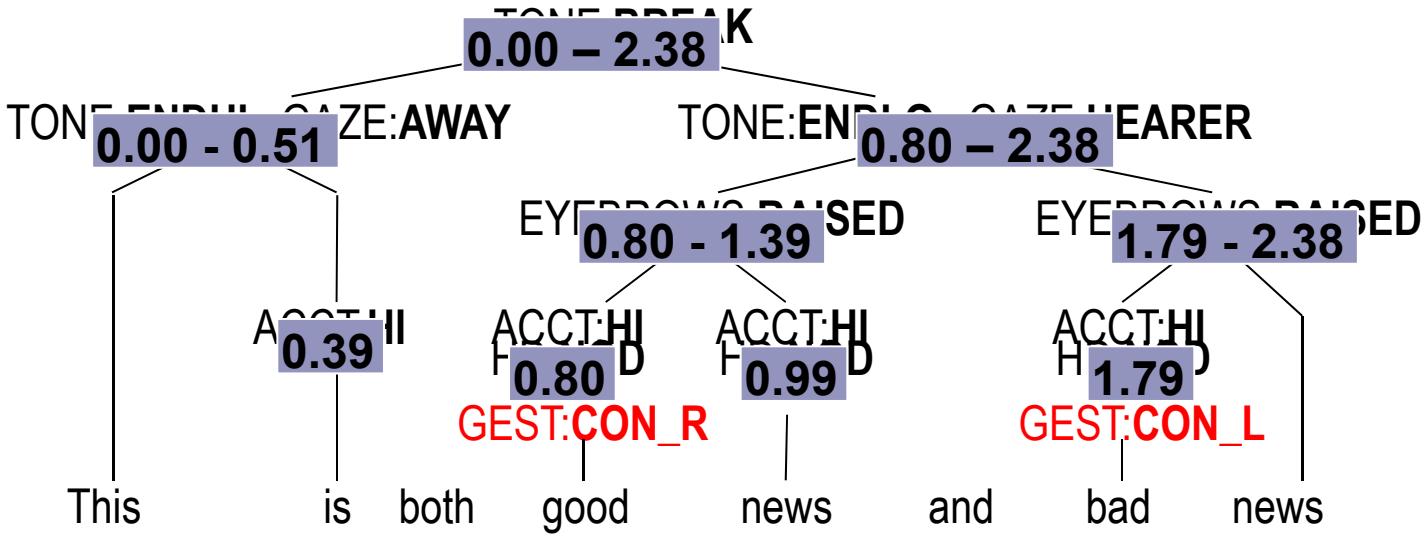
ジェスチャ生成 (ルールベースのアプローチ)

- 発話文へのジェスチャ付与
  - 文構造，シソーラス，非言語コミュニケーション研究の知見に基づき，非言語行動を決定[Cassell et al. 2001]
  - 文構造，コーパス調査に基づき，非言語行動を決定[Nakano et al. 2004]
- テキストに付与された非言語行動タグから実行スケジュールを作成
- スケジュールに従って音声とアニメーションを同期させて生成



ジェスチャ生成 (ルールベースのアプローチ)

- 発話文へのジェスチャ付与
  - 文構造，シソーラス，非言語コミュニケーション研究の知見に基づき，非言語行動を決定[Cassell et al. 2001]
  - 文構造，コーパス調査に基づき，非言語行動を決定[Nakano et al. 2004]
- テキストに付与された非言語行動タグから実行スケジュールを作成
- スケジュールに従って音声とアニメーションを同期させて生成



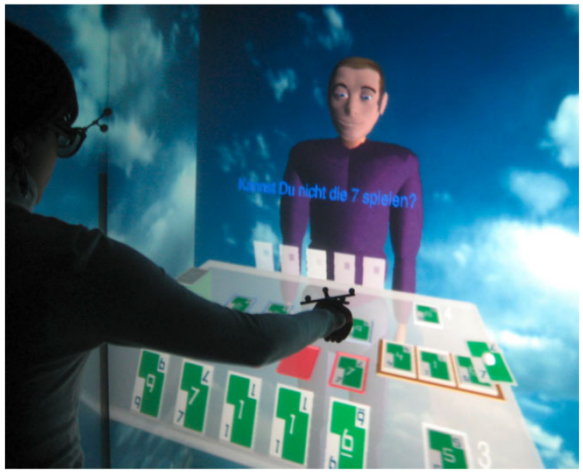
ジェスチャ生成(機械学習によるアプローチ)

- 訓練データ
  - 前フレームの動作
  - 韻律情報
    （ピッチ，インテンシティ等）
  - MFCC
  - 単語
  - POS
- 予測対象
  - ジェスチャタイミング(true/false)
  - ジェスチャラベル(Nクラス分類）
  - 次フレームの動き
- 系列学習による予測
  - CRF
  - RBM，LSTM

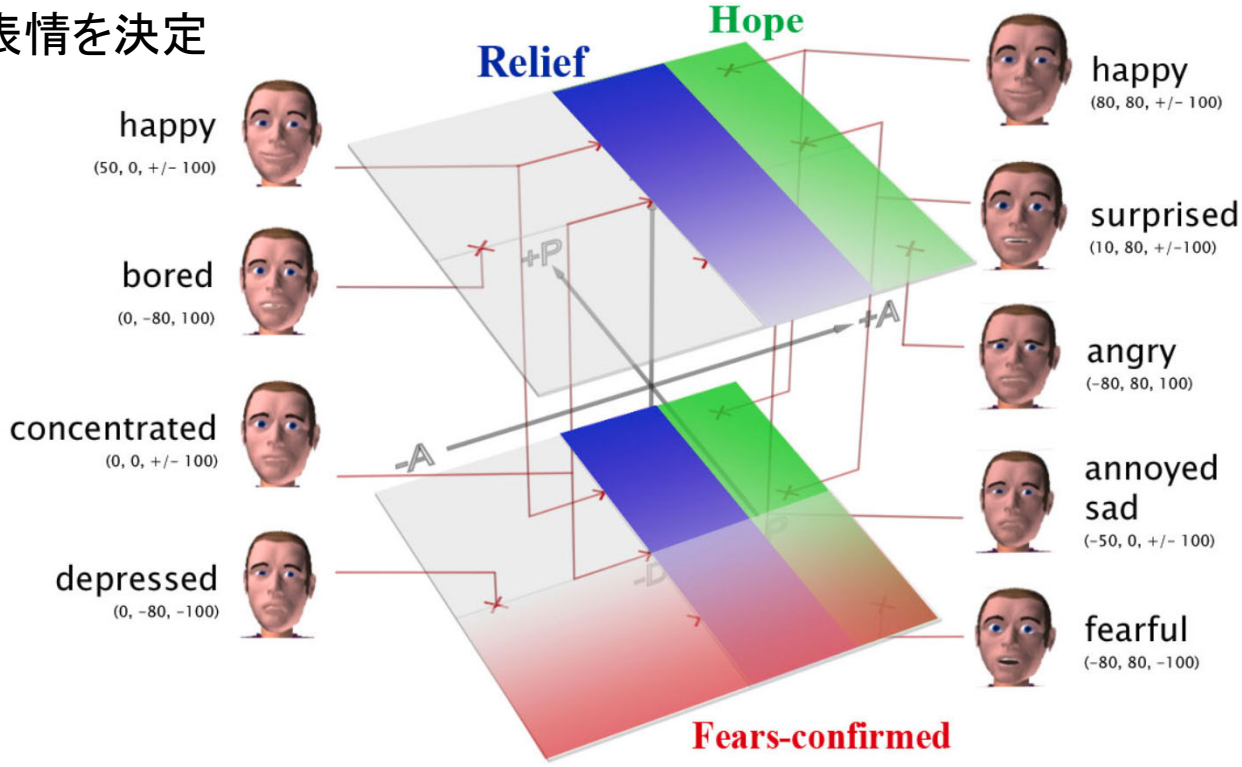
	訓練データ	予測対象	学習モデル
[Chiu et al. 2011]	前フレーム，韻律	次の動作フレーム	RBM
[Hasegawa et al. 2018]	MFCC	次の動作フレーム	LSTM
[Chiu et al. 2015]	韻律，単語，POS	ジェスチャラベル	深層学習＋系列学習
[[Ishii et al. 2018]	発話長，単語位置，BOW，対話行為，POS，シソーラス	ジェスチャ，頷き，表情，姿勢等	CRF

表情，感情表現の生成

- 感情モデルに基づく表情生成 [Becker-Asano et al. 2008]
  - 感情の3次元モデル：Valence，Arousal，Control/Dominance/Power
  - カードゲームの状態に応じてルールに基づき感情状態を更新
  - 感情モデルに基づき表情を決定



[Becker-Asano, 2008, Fig. 2]

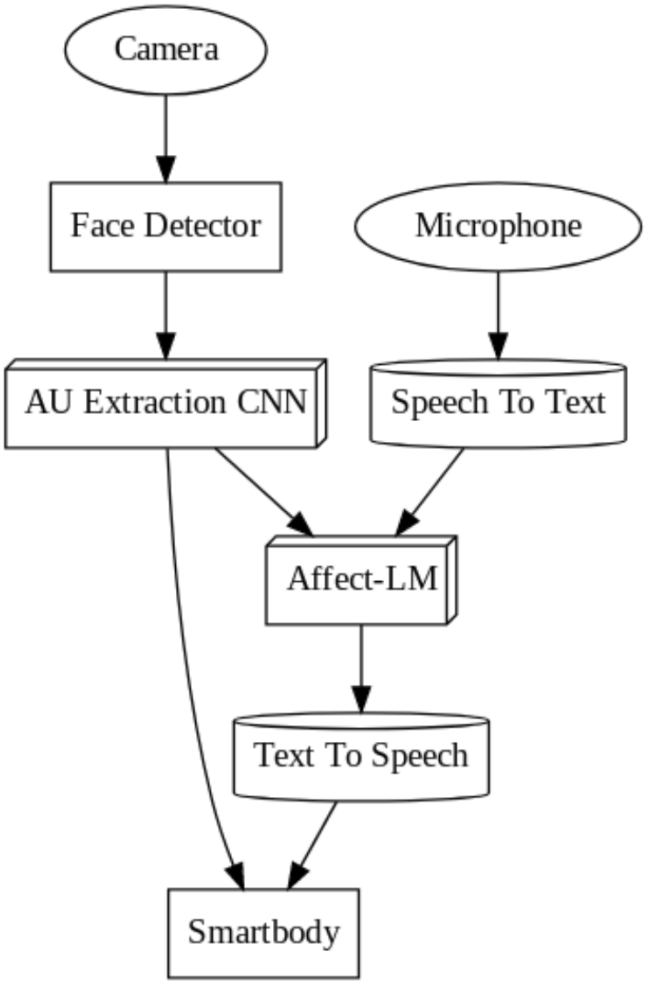


[Becker-Asano et al. 2008, Fig. 1]

Becker-Asano, C. and Wachsmuth, I. (2008). Affect Simulation with Primary and Secondary Emotions. IVA 2008, Tokyo.

感情生成

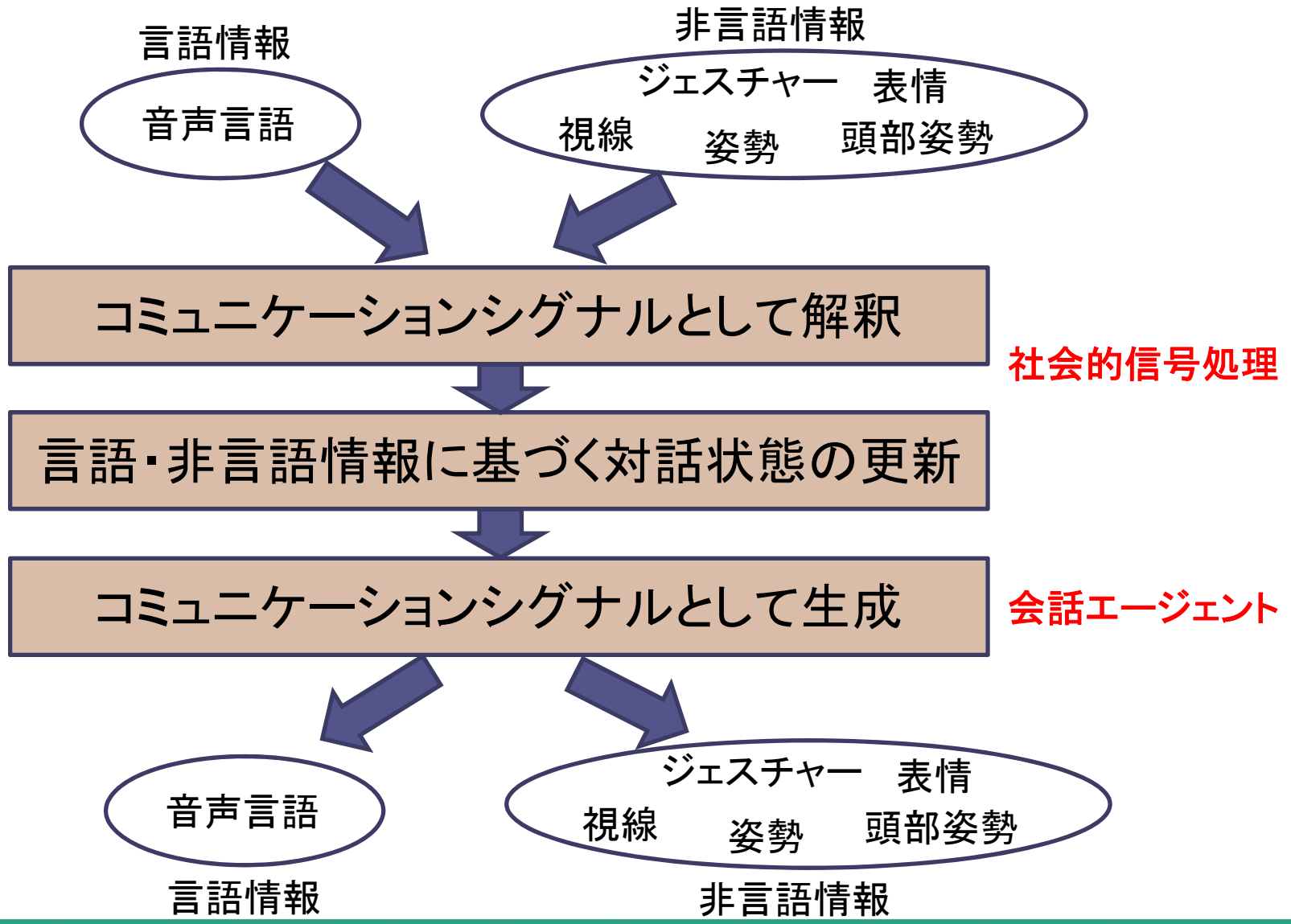
- Affect-LM [Ghosh et al. 2017]
  - 単語シーケンスと文に付与された5種類の感情タグを入力とするLSTMを学習
  - 5種類の感情表現文を生成
    I feel so...
- NaDiA [Wu et al. 2018]
  - 顔特徴検出器のデータからユーザの感情を推測
  - Affect-LMへの感情タグとして利用
  - ユーザ表情の模倣に利用



[Wu et al. 2018, Figure 1]

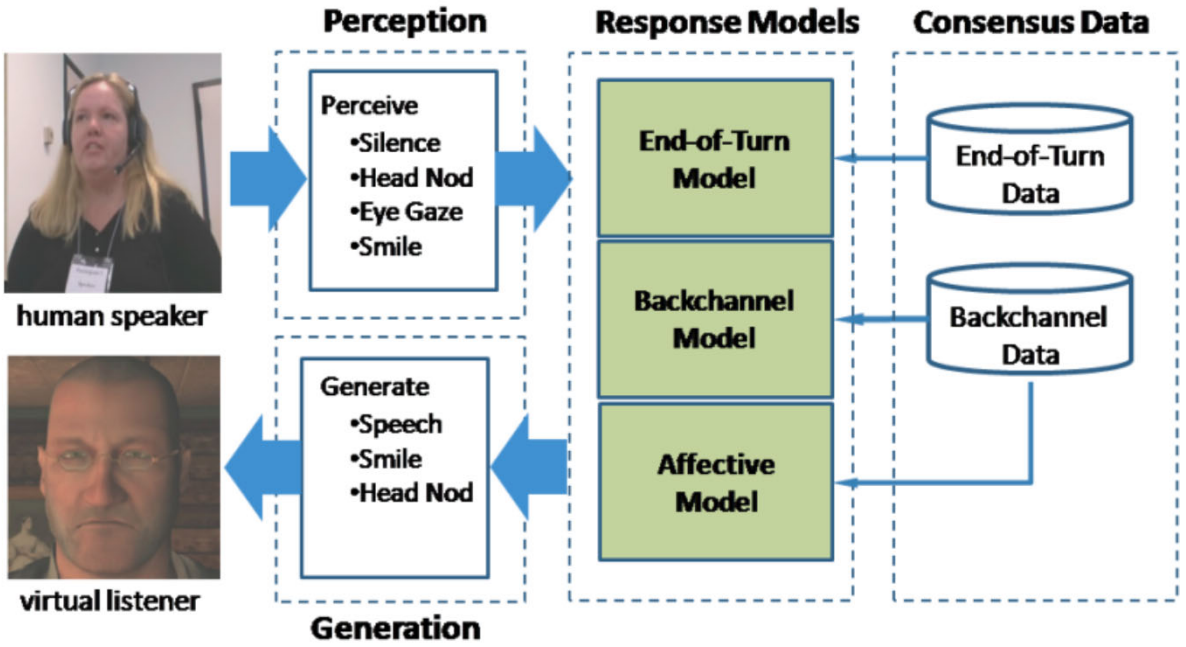
Wu, J., Ghosh, S., Chollet, M., Ly, S., Mozgai, S. and Scherer, S. (2018). NADiA - Towards Neural Network Driven Virtual Human Conversation Agents. the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '18), International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC.

統合的なシステム


言語情報
音声言語
非言語情報
ジェスチャー
表情
視線
姿勢
頭部姿勢
コミュニケーションシグナルとして解釈
社会的信号処理
言語・非言語情報に基づく対話状態の更新
コミュニケーションシグナルとして生成
会話エージェント
音声言語
言語情報
ジェスチャー
表情
視線
姿勢
頭部姿勢
非言語情報

視線，頷き等のフィードバック生成

- ラポール：会話相手と同調，調和していると感じること
- 南カリフォルニア大ICT　ラポールエージェント：非言語フィードバックによりラポールを確立するエージェント [Huang et al. 2011]
  - アイコンタクト，頷き，発話終了予測，笑顔の付与
- 会話が長続きする，自己開示が増える等の多くの効果が確認されている

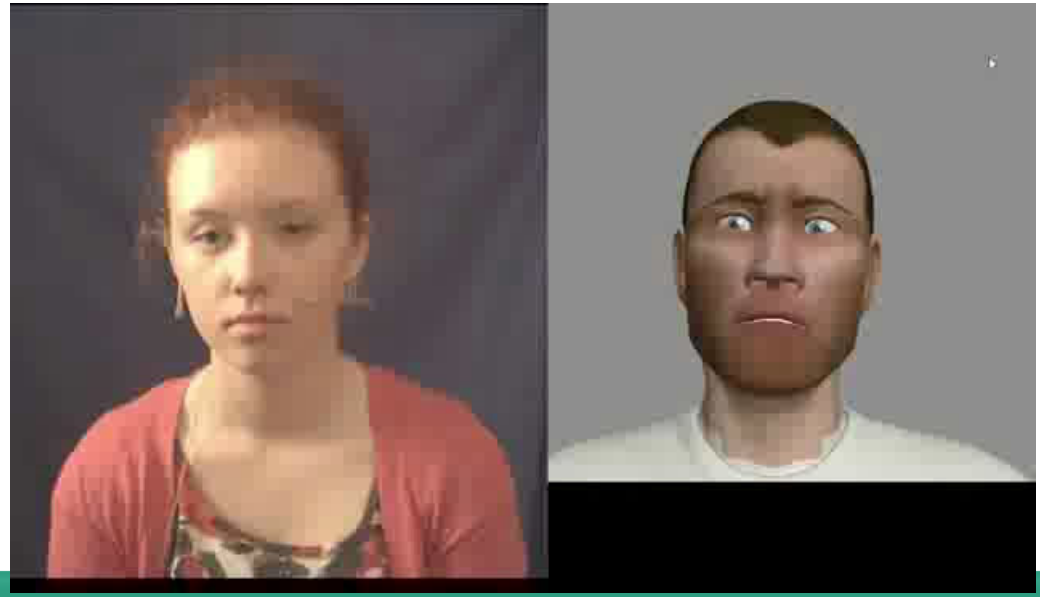


[Huang et al., 2011, Fig. 1]

Huang, L., Morency, L.-P. and Gratch, J. (2011). Virtual Rapport 2.0. IVA 2011.

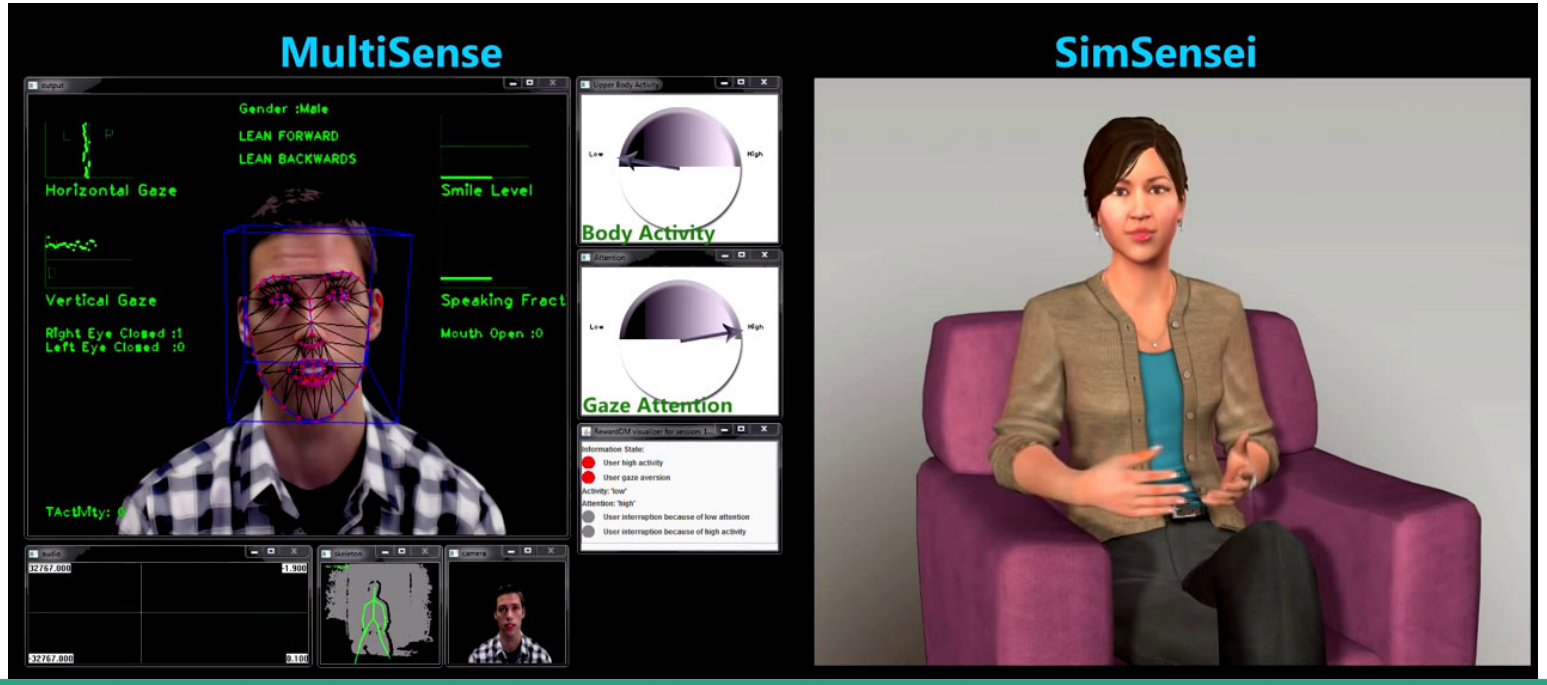
傾聴エージェント

- 傾聴エージェント：ユーザの話の聞き役となるエージェント
- （言語を理解することなく）会話を継続させる
- 非言語フィードバックの生成が中心課題
- SEMAINEプロジェクト(EU-FP7)
  - back-channelフィードバックを行う際の表情をコミュニケーションスタイルの違いに応じて決定
    - 肯定的なエージェント：笑顔で頷く
    - 威圧的(Aggressive)なエージェント：威圧的な表情，ユーザの発話を遮るフィードバック



インタビューエージェント

- 南カリフォルニア大ICT　SIMSENSEI [DeVault et al. 2014]
  - PTSDの診断支援のためのインタビューを行う
  - 非言語行動センシング：頭部動作，表情，視線，音声解析
  - 対話処理：音声認識，対話行為認識，発話感情(Pos/Neg)，対話制御
  - 生成機構：システム内部状態から行動スクリプト(BML)を生成，キャラクタアニメーション生成

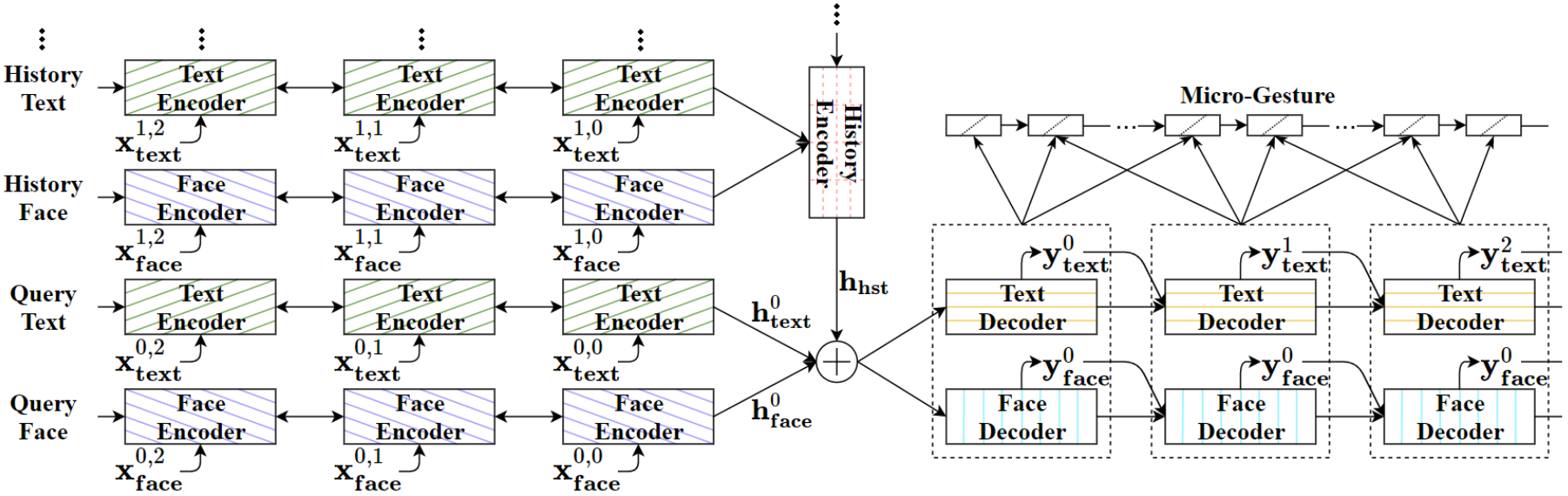


https://www.youtube.com/watch?v=ejczMs6b1Q4

Seq-to-Seqのマルチモーダル対話システム

- 表情＋頭部動作の表現
  - OpenFace により18種類のAUと3D頭部姿勢データを取得
  - ジェスチャテンプレート: 出現パターンをk-means(k=200)でクラスタリング
- 単語シーケンスとジェスチャテンプレートのシーケンスを統合した Seq-to-Seqモデルを作成
- デコーダ出力をTTSとアニメーションに変換してチャットボットを作成

デモ映像



[Chu et al. 2018, Figure 6]

http://www.cs.toronto.edu/face2face

Chu, H., Li, D. and Fidler, S. (2018). A Face-to-Face Neural Conversation Model. CVPR 2018.

インタラクション研究としてのマルチモーダル対話システム（まとめにかえて）

- 本日の内容
  - 社会的信号処理：コミュニケーションシグナルの理解
  - 会話エージェント：コミュニケーションシグナルの生成
  - マルチモーダル情報に基づく対話制御
- マルチモーダル対話システムは複雑
  - 多くの研究は理解か生成のどちらか
  - これらに対話制御も加えたマルチモーダル対話システムを実装するのは容易ではない
  - とはいえ，様々なツールや開発環境は整いつつある
- マルチモーダル性を理解・生成・対話制御のどこか1か所に加えるだけでも，マルチモーダル研究になる．
- インタラクションの観点から対話システムをとらえる→非言語も重要
  - 知的対話システム；正しく答える，かしこく答える
  - Human-Agent Interaction：ユーザがインタラクションにengageする，満足する，繰り返し使う

関連国際会議

- 社会的信号処理
  - ICMI
  - FG
  - ACII
- 会話エージェント
  - IVA
  - AAMAS
  - HRI
  - HAI
- 対話システム
  - SIGDIAL
  - ACL

参考文献

- Aran, O. and Gatica-Perez, D. (2013). One of a Kind: Inferring Personality Impressions in Meetings. the 15th ACM on International conference on multimodal interaction (ICMI2013)**:** 11-18.
- Bavelas, J. B. (1994). "Gestures as Part of Speech: Methodological Implications." Research on Language and Social Interaction **27**(3): 201-221.
- Becker-Asano, C. and Wachsmuth, I. (2008). Affect Simulation with Primary and Secondary Emotions. IVA 2008, Tokyo.
- Bohus, D. and Horvitz, E. (2009). Learning to Predict Engagement with a Spoken Dialog System in Open-World Settings. SIGdial'09, London, UK.
- Bohus, D. and Horvitz, E. (2010). Facilitating Multiparty Dialog with Gaze, Gesture, and Speech. The International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI'10).
- Carletta, J., et al., P. (2005). The AMI meeting corpus: a pre-announcement. the Second international conference on Machine Learning for Multimodal Interaction (MLMI'05), Springer-Verlag, Berlin, Heidelberg.
- Cassell, J., Vilhjalmsson, H. and Bickmore, T. (2001). BEAT: The Behavior Expression Animation Toolkit. SIGGRAPH 01, Los Angeles, CA, ACM Computer Graphics Press.
- Chiu, C.-C. and Marsella, S. (2011). How to Train Your Avatar: A Data Driven Approach to Gesture Generation. Intelligent Virtual Agents. IVA 2011, Springer, Berlin, Heidelberg.
- Chiu, C.-C., Morency, L.-P. and Marsella, S. (2015). Predicting Co-verbal Gestures: A Deep and Temporal Modeling Approach. Intelligent Virtual Agents. IVA 2015, Springer.
- Chu, H., Li, D. and Fidler, S. (2018). A Face-to-Face Neural Conversation Model. CVPR 2018.

参考文献(Cont.)

- DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., Lucas, G., Marsella, S., Morbini, F., Nazarian, A., Scherer, S., Stratou, G., Suri, A., Traum, D., Wood, R., Xu, Y., Rizzo, A. and Morency., L.-P. (2014). SimSensei kiosk: a virtual human interviewer for healthcare decision support. the 2014 international conference on Autonomous agents and multi-agent systems (AAMAS '14).
- Fang, S., Achard, C. and Dubuisson, S. (2016). Personality classification and behaviour interpretation: an approach based on feature categories. the 18th ACM International Conference on Multimodal Interaction (ICMI 2016).
- Gale Lucas, G. S., Shari Lieblich, and Jonathan Gratch (2016). Trust me: multimodal signals of trustworthiness. the 18th ACM International Conference on Multimodal Interaction (ICMI 2016), ACM, New York, NY, USA.
- Ghosh, S., Chollet, M., Laksana, E., Morency, L.-P. and Scherer, S. (2017). Affect-LM: A Neural Language Model for Customizable Affective Text Generation. the 55th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics
- Hasegawa, D., Kaneko, N., Shirakawa, S., Sakuta, H. and Sumi, K. (2018). Evaluation of Speech-to-Gesture Generation Using Bi-Directional LSTM Network. 18th ACM International Conference on Intelligent Virtual Agents (IVA2018).
- Huang, L., Morency, L.-P. and Gratch, J. (2011). Virtual Rapport 2.0. IVA 2011.
- Ishii, R., Katayama, T., Higashinaka, R. and Tomita, J. (2018). Generating Body Motions using Spoken Language in Dialogue. 18th ACM International Conference on Intelligent Virtual Agents (IVA2018).
- Ishii, R., Otsuka, K., Kumano, S. and Yamato, J. (2014). Analysis of Respiration for Prediction of "Who Will Be Next Speaker and When?" in Multi-Party Meetings. the 16th International Conference on Multimodal Interaction (ICMI '14).

参考文献(Cont.)

- Ishii, R., Otsuka, K., Kumano, S. and Yamato, J. (2016). "Prediction of Who Will Be the Next Speaker and When Using Gaze Behavior in Multiparty Meetings." ACM Trans. Interact. Intell. Syst.: 2160-64552160-64556455.
- Kumano, S., Otsuka, K., Mikami, D., Matsuda, M. and Yamato, J. (2012). Understanding communicative emotions from collective external observations. CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12), ACM, New York, NY, USA.
- Nakano, Y. I. and Ishii, R. (2010). Estimating User's Engagement from Eye-gaze Behaviors in Human-Agent Conversations. 2010 International Conference on Intelligent User Interfaces (IUI2010), Hong Kong.
- Nakano, Y. I., Okamoto, M., Kawahara, D., Li, Q. and Nishida, T. (2004). Converting Text into Agent Animations: Assigning Gestures to Text. Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2004), Companion Volume, Boston.
- Nguyen, L. S., Marcos-Ramiro, A., Romera, M. M. and Gatica-Perez, D. (2013). Multimodal analysis of body communication cues in employment interviews. Proceedings of the 15th ACM on International conference on multimodal interaction. Sydney, Australia, ACM**:** 437-444.
- Nguyen, L. S., Frauendorfer, D., Mast, M. S. and Gatica-Perez, D. (2014). "Hire me: Computational inference of hirability inemployment interviews based on nonverbal behavior." IEEE Transactions on Multimedia **16**(4): 1018-1031.
- Nojavanasghari, B., Gopinath, D., Koushik, J., Baltru, T., #353, aitis and Morency, L.-P. (2016). Deep multimodal fusion for persuasiveness prediction. Proceedings of the 18th ACM International Conference on Multimodal Interaction. Tokyo, Japan, ACM**:** 284-288.
- Okada, S., Aran, O. and Gatica-Perez, D. (2015). Personality Trait Classification via Co-Occurrent Multiparty Multimodal Event Discovery. the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15).

参考文献(Cont.)

- Okada, S., Nguyen, L. S., Aran, O. and Gatica-Perez, D. (2018). "Modeling Dyadic and Group Impressions with Inter-Modal and Inter-Person Features." ACM Transactions on Multimedia Computing, Communications, and Applications, **9**(4): Article 39.
- 岡田将吾, 松儀良広, 中野有紀子, 林佑樹, 黄宏軒, 高瀬裕,新田克己 (2016). "マルチモーダル情報に基づくグループ会話におけるコミュニケーション能力の推定." 人工知能学会論文誌 **31**(6): AI30-E_31-12.
- Pentland, A. (2008). Honest Signals: How They Shape Our World, The MIT Press.
- Poria, S., Cambria, E., Hazarika, D., Majumder, N., Zadeh, A. and Morency, L.-P. (2017). Context-Dependent Sentiment Analysis in User-Generated Videos. the 55th Annual Meeting of the Association for Computational Linguistics.
- Roddy, M., Skantze, G. and Harte, N. (2018). Multimodal Continuous Turn-Taking Prediction Using Multiscale RNNs. the 20th International Conference on Multimodal Interaction (ICMI '18).
- Sanchez-Cortes, D., Aran, O., Jayagopi, D. B., Mast, M. S. and Gatica-Perez, D. (2013). "Emergent leaders through looking and speaking: from audio-visual data to multimodal recognition." Journal on Multimodal User Interfaces **7**(1-2): 39-53.
- Skantze, G. (2017). Towards a General, Continuous Model of Turn-Taking in Spoken Dialogue Using LSTM Recurrent Neural Networks. SigDial, Saarbrucken, Germany.
- Vinciarelli, A., Pantic, M. and Bourlard, H. (2009). "ocial signal processing: Survey of an emerging domain." Image and Vision Computing **27**: 1743–1759.
- Wu, J., Ghosh, S., Chollet, M., Ly, S., Mozgai, S. and Scherer, S. (2018). NADiA - Towards Neural Network Driven Virtual Human Conversation Agents. the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '18), International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC.

レビュー論文

- Vinciarelli, A., Pantic, M. and Bourlard, H. (2009). "ocial signal processing: Survey of an emerging domain." Image and Vision Computing **27**: 1743–1759.
- Baltrusaitis, T., Ahuja, C. and Morency, L.-P. (2018). "Multimodal Machine Learning: A Survey and Taxonomy." IEEE transactions on pattern analysis and machine intelligence.
- 岡田将吾, 石井亮 (2017). "社会的信号処理とAI (特集 2017年度人工知能学会全国大会(第31回))." 人工知能 : 人工知能学会誌 : journal of the Japanese Society for Artificial Intelligence 32(6): 915-920.