マルチモーダル対話データの収集と興味判定アノテーションの分析

Collection of Multimodal Dialog Data and Analysis of the Result of Annotation of Users' Interests

荒木雅弘[1*]　冨増紗也華[1]　中野幹生[2]　駒谷和範[3]
Masahiro Araki[1]　Sayaka Tomimasu[1]　Mikio Nakano[2]　Kazunori Komatani[3]

岡田将吾[4]　藤江真也[5]　杉山弘晃[6]
Shogo Okada[4]　Shinya Fujie[5]　Hiroaki Sugiyama[6]

[1] 京都工芸繊維大学　[2] HRI-JP　[3] 大阪大学
[1] Kyoto Institute of Technology　[2] HRI-JP　[3] Osaka University
[4] 北陸先端科学技術大学院大学　[5] 千葉工業大学　[6] NTT
[4] JAIST　[5] Chiba Institute of Technology　[6] NTT

Abstract: Human-System Multimodal Dialogue Sharing Corpus Building Group is acting as a working group of SIG-SLUD for the purpose of constructing corpus for evaluating elemental technologies of the multimodal dialogue system. In this paper, we report on the results of recording chat dialogue data between human subjects and virtual agents by the WoZ method conducted in 2016 and the result of the analysis of annotations of the users' interests in the data.

1　はじめに

言語処理・音声処理・画像処理やそれを支える機械学習技術の発展と，計算機およびセンサ・ディスプレイ・ロボットなどのハードウェア技術の進歩により，音声言語のみならず画像その他のモダリティを用いて人間とコミュニケーションを行うマルチモーダル対話システムの研究が盛んになっている．しかしながら，人間とコミュニケーションを行う際に，システムがどのようにマルチモーダル情報を利用すべきかはまだ十分明らかになっているとは言い難い．この理由の一つとして，ユーザの意図・感情・態度や対話の状況に関する情報などがアノテーションされたマルチモーダル対話コーパスが整備・共有されていないことがあげられる．一方で，マルチモーダル対話データに対してアノテーションを行うことは，多大なコストがかかるため，複数の研究機関が協力して作業できるのが望ましい．

このような背景のもと，人工知能学会 言語･音声理解と対話処理研究会のワーキンググループとして，「人システム間マルチモーダル対話共有コーパス構築グループ」の活動が 2016 年 4 月に開始された．

*連絡先：京都工芸繊維大学
京都市左京区松ヶ崎御所海道町
E-mail: araki@kit.ac.jp

人の会話のマルチモーダル分析やマルチモーダルコーパスの共有は，複数の人どうしの会話を対象としたものが数多く行われている（[1], [2]）．これに対して，本プロジェクトでの対象は，人対人ではなく，人対システムの対話である．ユーザがシステムを相手にどうふるまうかというデータを収集し共有することで，マルチモーダル対話システム研究の要素技術の開発に資することを目指している．

本稿では，これまでワーキンググループで収集したマルチモーダル対話データに関して，収集法（2 章）およびアノテーション内容（3 章）を報告し，アノテーション結果の分析（4 章）を通じて，今後のデータ収集・アノテーション方法について議論する．また，関連研究との違いを踏まえつつ（5 章），今後のデータの一般公開に向けた問題点を整理して報告する（6 章）．

2　人システム間マルチモーダル雑談対話データの収集

対話エージェント（画面上に人型のキャラクタを表示してユーザと対話するシステム）と人間との対話において，人間がどのように振る舞うかを分析するため

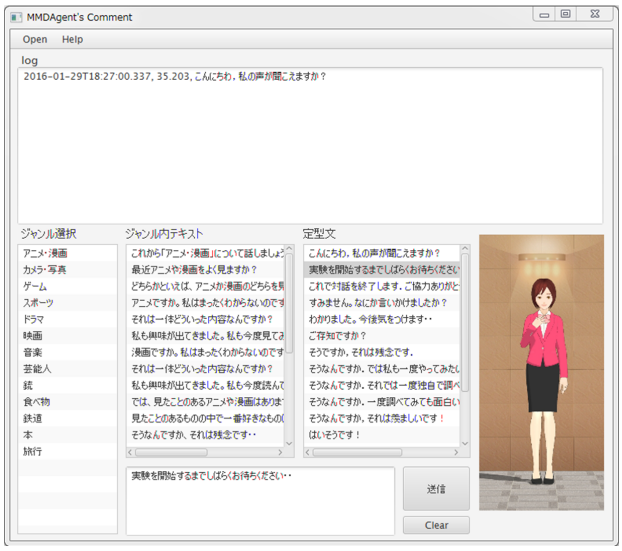

図 1: 操作者側のインタフェース

のデータ収集環境を構築した [19]．対話エージェントは MMDAgent [1]を用い，遠隔から操作する WoZ (Wizard of Oz) 方式で制御する．

操作者側は，図 1 に示すインタフェースを用いて現在の話題を選択すると，その話題に特化した発話例がリストアップされる．また，どの話題でも共通して使うことができる発話例は常時表示されており，操作者はいずれかの発話を選んで送信ボタンを押すと，エージェントの発話がおこなわれる．また一部の発話の末尾には，喜び・疑問・微笑み等を現す特定の文字列が付与されており，発話に伴うエージェントの表情変化や動作を制御している．

データ収集に際して，操作者には，被験者の発話の積極性に応じて，発話を促す質問や受け身的な応答の割合を適度に変えることで，話題ごとの交換数がほぼ一定になるようにインストラクションを与えた．

また被験者には，準備した 10 数個の話題について，事前に興味あり／なしをマークしてもらい，興味あり／なしとマークされた各 3 話題について，適当な順序で雑談の話題とした．データ収録では，各話題についてシステム側から話を切り出し，事前に準備した質問や想定される応答に対する反応などを返しながら，10 数回の交換をおこなった．被験者の上半身正面画像と音声は全データについて，また一部のデータについて Microsoft 社製 Kinect V2 によるセンサー情報の記録をおこなった．図 2 に対話例を示す．

2016 年 4 月に試験的に 4 名分のデータ（以降，データ 1）を収録し，収録環境における問題点を検討した．

[1]http://www.mmdagent.jp/

S:これから鉄道 (0.0) について話しましょう．
U:はい ((snuff))
S:電車は好きです↑か↓：
U:まあ別に (0.0) 好きでも嫌いでもないですね
S:では，電車はよく利用しますか？
U:ん，まああんまり最近乗らないですね，((snuff))
S:どういった時に電車を利用するんですか
U:なんか，ん :: まあちょっと :: (0.0) 県外というか府外に出かけるとか :，(0.0) まあお酒飲む (.) からあんまり>原付き<とか使えへんなあっていうとき (0.2) ですね

図 2: 雑談対話例

この際に，ワーキンググループメンバーまたは関係者による興味判定アノテーションを検討のためにおこなった．話題提示の順序をランダムにすることや，被験者へのインストラクションを明確にすることなどの改善をおこなって，2017 年 1 月に新たに 10 名分のデータ（以降，データ 2）を収録した．データ 2 に関しては，予備アノテーションを 3 名のアノテータによっておこない，主としてアノテーションマニュアルに関する改善をおこなった後に，別機関で本アノテーションを 3 名のアノテータによっておこなった．

アノテーションにおいては，人と対話システムとの雑談対話において，人がその対話内容に興味を持っているのかどうかを，1 交換（「システムの発話 S」－「ユーザの発話 U」の対）毎に表情・音声の韻律情報・発話内容などから判断し，興味あり（o），不明（t），興味なし（x）のラベルを付与した．

3 アノテーションスキーマの確立に向けて

システムとの雑談対話において，ユーザがその話題に興味を持っているのか否かが判定できれば，システムがユーザに関するプロファイルを把握することができ，より個人に適応した対話システムを実現することができる．ここでは，そのような興味の有無のアノテーションに関して，今後の分析や学習用データとして信頼できる結果が得られるかどうか，そのためのアノテーション指針はどのようにすればよいかについて検討をおこなった．

具体的には，データ 2（10 人分，789 交換）を題材として，まず，人と機械とのインタラクションを研究している大学院生 3 名によって非常に直観的なインストラクションによるアノテーションをおこなった．判定が分かれたデータに対する考察を通じてアノテーションマニュアルを整備し，別のアノテータ 3 名によるアノテーション結果と比較した．

表 1: アノテーションの結果

ラベル	1 回目	2 回目
興味あり	907	992
不明	162	267
興味なし	1276	1108
エラー	22	0
Fleiss' kappa	0.407	0.490

最初のアノテーションに関しては，動画データを 1 交換毎に再生し，被験者がその話題に興味を持っているか否かを直観的に判定した．ただし，話を楽しんでいるかどうかの判定ではなく，あくまでも現在の話題に対する興味の有無を判定することとした．結果を表 1 の「1 回目」の列に示す．このアノテーションでは，被験者からの単純な聞き返しはエラーと判定している．

収録においては，事前の調査に基づいて，興味がある話題とない話題の数が同数になるようにコントロールしているので，興味あり／なしのラベルが同数程度であることが望ましいが，興味なしと判定された交換がやや多くなっている．Fleiss の κ 値は 0.407 であり，解釈としては Fair agreement と Moderate agreement の境界程度の値である．

このデータを元に，アノテータ 3 名が特に判定が分かれた交換について，それぞれどのような解釈で判定したかを議論した．その結果，以下に示すように判定が分かれた原因がいくつか抽出された．

- 話題の冒頭部分に関して，まだほとんど情報がないと考えるアノテータと，他の交換と区別せずに少ない情報からでもできるだけどちらかに判定しようとするアノテータがいた．
- 交換の全体を見て判定するアノテータと，交換の一部の特徴的なところを捉えて判定するアノテータがいた．
- 特定のアノテータが特定のモダリティ（たとえば笑顔）を常に優先的な情報としている場合があった．
- 興味の有無の現れ方が，被験者間共通の絶対的なものか，被験者内で相対的なものかについて解釈のぶれがあった．

以上の点を考慮し，以下のようなインストラクションを記述したアノテーションマニュアルを作成して，別の研究機関勤務のアノテータ 3 名でアノテーションをおこなった．

- 話題の冒頭部分は，積極的な理由がない限り「不明」とする．
- 交換の全体から判定する．
- 特定のモダリティに偏った判定はしない．
- 1 人の被験者について最初にデータを通して見て，被験者の感情表出の大きさやくせなどを把握する．
- エラーは完全に実験想定外の発話に限定する．

結果を表 1 の「2 回目」の列に示す．Fleiss の κ 値は 0.490 となり，解釈としては Moderate agreement といえるレベルまで一致率が上がった．

本実験の結果，人とシステムの対話という特殊な状況下かつ個人差が大きいデータを対象にした興味の有無の判定という難しい問題について，適切なインストラクションをおこなうことによってある程度信頼できるアノテーションが得られることがわかった．

4 予備アノテーション（データ 1）結果におけるラベルの偏りの分析

対話への興味度のアノテーションはアノテータの主観に依存すると考えられる．このため，各アノテータのアノテーション傾向を知ることが，予備アノテーションにおける一つの主要な目的であった．本節では，予備アノテーションデータ（データ 1）における各アノテータの偏りの分析を行う．4 人の被験者とシステムとの対話を含むデータ 1 に対して，8 人のアノテータが各交換毎に「興味あり」,「ニュートラル」,「興味なし」の 3 つのいずれかのラベルを付与した．

8 人のアノテータ間の一致率を測るために Fleiss の κ 値 （κ_f）を算出した結果，3 ラベルに対する一致率は 0.26 であった．この値はランダムよりは高いが，絶対値としては低い値であり，各アノテータのアノテーションに偏りが生じていることを示す．本節では，偏りを（1）アノテータ間でのアノテーション傾向の違い，（2）一致率（κ_f）の被験者間の差の観点から調査する．

4.1 アノテータ間でのアノテーション傾向の分析

8 人のアノテータ間（A1-A8）の各ペアで Cohen の κ 値 （κ_c）を計算し，距離尺度に変換（1-κ_c）した後に Ward 法で階層クラスタリングを行った結果を図 4 に示す．また，各ペアの κ_c を図 3 に示す．値が大きいほど対象箇所を濃い色で示している．

あるアノテータ X と，他の 7 人のアノテータ群との一致度を，X と 7 人との κ_c の平均値（κ_{Av}）と定義し，これを付加情報としてクラスタリング結果を解釈する．

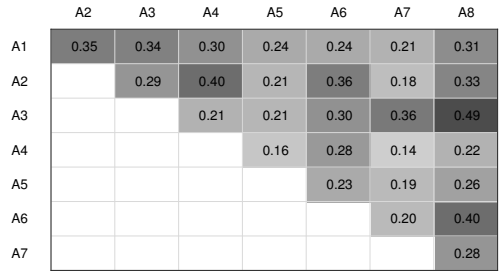

図 3: 各アノテータ間の Cohen's Kappa 値

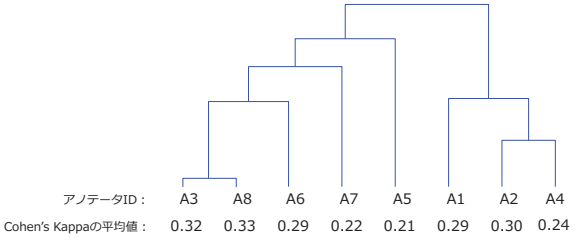

図 4: アノテータのクラスタリング結果

κ_{Av} は X のアノテーションが他の 7 人とどの程度一致していたかを示す指標である．図 4 より，κ_{Av} が最も低い，二番目に低いアノテータは，それぞれ A5(0.213), A7 (0.222) である．またクラスタ数が 4 の場合，A5, A7 はどのアノテータともマージされず単一サンプルのクラスタを形成している．以上の理由より，今回のアノテーションタスクにおいて，A5，A7 は他のアノテータとは異なる傾向でアノテーションを行っていると考えられる．アノテータのクラスタリングを通じて，アノテーション傾向を分析した結果，このアノテーションタスクが個人の主観性に影響を受けること，また特定のアノテータ間のアノテーションの相違・類似傾向が明らかになった．

4.2 各被験者に対するアノテーション一致率の分析

対話システムへの回答方法,「興味」を示す際に表出する非言語情報の種類・多寡は，4 人の被験者ごとに異なるため，被験者ごとに「興味度」のアノテーションの一致率も異なると考えられる．被験者ごと（P1-P4）で一致率を計算した結果を図 5 に示す．黒色の棒グラフは,「興味あり」,「ニュートラル」,「興味なし」の 3 値ラベルに対するアノテーションの一致率を，灰色の棒グラフは「ニュートラル」と「興味なし」を一つのラベルにマージし構築した 2 値ラベルに対するアノテーションの一致率を示す．P1 と P2 の 3 値，2 値ラベルに

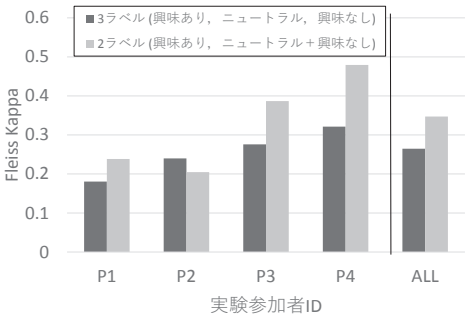

図 5: 被験者ごとのアノテーション一致度

対する一致度は，P3，P4 のそれより低い．3 値ラベルに対する一致度の最大値は P4 で 0.32，最小値は P1 で 0.18，2 値ラベル対する一致度の最大値も P4 で 0.48, 最小値は P2 で 0.20 であった．P1 や P2 のように，アノテーションが一致しない被験者の印象推定は，システムにとっても難しいタスクであることが予想されるため，興味度を判定することが困難である発言や被験者をシステムによって特定する方法を構築することも，有用な将来課題であると考えられる．

5 関連研究

人の会話のマルチモーダル分析やマルチモーダルコーパスの共有を目的として，海外では，複数人が参加する会議データが，AMI (Augmented Multi-party Interaction) [1] や ICSI meeting corpus [2] として公開されている．また，CHIL (Computers in Human Interaction Loop) [3] ではオフィスや教室でなされるインタラクションが，VACE (Video Analysis and Content Extraction) [4] では空軍における戦闘ゲームセッションにおけるインタラクションが，それぞれ対象とされている．

人どうしの対話のマルチモーダル分析の研究も多く行われている．例えば，カウンセリング対象の状態を推定し，意思決定支援に用いることができるシステムが有名である [5]．国内でも人間どうしの会話における視線の自動推定 [6] や，漫才におけるマルチモーダル分析の研究がある [7, 8, 9]．

これらに対して本プロジェクトでは，マルチモーダル対話システム研究の要素技術の開発に資することを目指して，人対システムの対話を対象としている．対話のできるロボットの実現は，対話システム研究のゴールのひとつであり，多くの研究がなされている [10, 11, 12, 13, 14, 15]．また，対話においてユーザの興味を検出する研究も行われており [16, 17, 18, 19]，その検出結果に応じて，話題を続けたり話題を深めたりといった

ように，その後の対話を変化させることができる．本プロジェクトは，これらの研究の延長線上に位置し，人対システムの対話において，ユーザの興味を含むマルチモーダルなデータを共有し，これらの研究の基盤となることを目指している．人対人のデータと，人対システムのデータの最大の違いは，相手がシステムであることをユーザが意識しているか否かである．ユーザは，人に対する場合とシステムに対する場合とで，異なったふるまいをする．人対システムのデータを収集することで，実際にシステムを構築した際に，システムに対してユーザが行うであろうふるまいが収集できる．また，対話システムでは，言語を用いたやりとりが複数ターンにわたって続き，かつ，システムは対話状態を持つという特徴がある．現在データ収集に用いている WoZ システムでも，システムから提示する話題には明らかな区切りがあり，これは対話状態のひとつである．単なるマルチモーダルデータではなく，このような対話状態を持つシステムとの対話データを収集することで，一問一答的な対話ではなく，対話の進行を考慮したシステム設計などに繋がる可能性がある．

人対システムの対話コーパスの共有という視点では，テキストでの雑談対話の収集や共有，shared task の実施などの取組みが国内でも行われている [20]．また現在，音声入力のチャットボットのコンテストとして，Amazon Alexa Challenge[2]も行われている．本プロジェクトは，テキストや音声入力だけでなく，マルチモーダルな対話システムが対象である．

また本プロジェクトでは，上述の雑談対話コーパス [20] と同様に，複数人によるアノテーション結果を付与して公開することを予定している．付与対象である興味は，アノテータの主観に基づき付与されるため，一意に定められる正解を付与するという問題ではなく，個人差が不可避である．このような主観に基づくアノテーション結果をどう扱い，対話システムの研究にどう生かすかといった研究にも利用可能である．

6　公開へ向けての課題

現在，総務省の方針として，個人識別性を有するデータは，パーソナルデータとしてその利用・流通に関してガイドラインが定められている [21]．今回収録したデータは本人の顔画像を含んでおり，パーソナルデータに該当すると考えられる．パーソナルデータには，主としてプライバシー保護の観点から，いくつかの利活用の原則が定められているが，研究用の共有データとする場合には，以下の問題点を事前に解決しておく必要がある．

- 被験者に対して，データの利用範囲についての同意を求める必要があるため，事前にどのような研究をおこなうかを慎重に検討しておく必要がある．
- パーソナルデータの本人は，当該パーソナルデータの取扱いについて同意した場合であっても当該同意を撤回することが可能であることが望ましいとされている．そのため，研究用データの配布を業とする機関が配布元となる必要がある．
- このような性質上，データのアノテーション等に関しても，クラウドソーシングなどの方法が使えず，良質な研究用データとするためには継続的に関与する研究グループを維持する必要がある．

現在，ワーキンググループ内では，データの公開に向けて上記問題に対する検討をおこなっているところである．

謝辞

ワーキンググループの活動を支援してくださる言語・音声理解と対話処理研究会主査 伝康晴氏に深く感謝いたします．また，データ収集・アノテーションにご協力いただいたワーキンググループのメンバーおよびデータ公開に関する助言をいただいた国立情報学研究所 大須賀智子氏に感謝いたします．

参考文献

[1] Carletta, J.: Unleashing the killer corpus: experiences in creating the multi-everything AMI Meeting Corpus, *Language Resources and Evaluation*, Vol. 41, No. 2, pp. 181–190 (2007).

[2] Janin, A., Baron, D., Edwards, J., Ellis, D., Gelbart, D., Morgan, N., Peskin, B., Pfau, T., Shriberg, E., Stolcke, A. and Wooters, C.: The ICSI Meeting Corpus, *Proc. IEEE International Conference on Acoustics, Speech & Signal Processing (ICASSP)*, pp. I–364–I–367 (2003).

[3] Waibel, A. and Stiefelhagen, R.: *Computers in the Human Interaction Loop*, Springer Publishing Company, Incorporated, 1st edition (2009).

[4] Chen, L., Rose, R. T., Qiao, Y., Kimbara, I., Parrill, F., Welji, H., Han, T. X., Tu, J., Huang, Z., Harper, M., Quek, F., Xiong, Y., McNeill,

[2]https://developer.amazon.com/alexaprize

D., Tuttle, R. and Huang, T.: VACE Multimodal Meeting Corpus, *Proceedings of the Second International Conference on Machine Learning for Multimodal Interaction (MLMI05)*, pp. 40–51 (2006).

[5] Stratou, G. and Morency, L.-P.: MultiSense—Context-Aware Nonverbal Behavior Analysis Framework: A Psychological Distress Use Case, *IEEE Transactions on Affective Computing*, Vol. 8, No. 2, pp. 190–203 (2017).

[6] 大塚和弘, 竹前嘉修, 大和淳司, 村瀬洋: 複数人物の対面会話を対象としたマルコフ切替えモデルに基づく会話構造の確率的推論, 情報処理学会論文誌, Vol. 47, No. 7, pp. 2317–2334 (2006).

[7] 林宏太郎, 神田崇行, 宮下敬宏, 石黒浩, 萩田紀博: ロボット漫才: 社会的受動メディアとしての二体のロボットの利用, 日本ロボット学会誌, Vol. 25, No. 3, pp. 381–389 (2007).

[8] 川嶋宏彰, スコギンズ・リーバイ, 松山隆司: 漫才の動的構造の分析：間の合った発話タイミング制御を目指して, ヒューマンインタフェース学会論文誌, Vol. 9, No. 3, pp. 379–390 (2007).

[9] 岡本雅史, 大庭真人, 榎本美香, 飯田仁: 対話型教示エージェントモデル構築に向けた漫才対話のマルチモーダル分析, 知能と情報, Vol. 20, No. 4, pp. 526–539 (2008).

[10] 松坂要佐, 東條剛史, 小林哲則: グループ会話に参与する対話ロボットの構築, 電子情報通信学会論文誌, Vol. J84-D-II, No. 6, pp. 898–908 (2001).

[11] Bohus, D. and Horvitz, E.: Models for Multiparty Engagement in Open-World Dialog, *Proc. Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, pp. 225–234 (2009).

[12] Al Moubayed, S., Skantze, G., Beskow, J., Stefanov, K. and Gustafson, J.: Multimodal Multiparty Social Interaction with the Furhat Head, *Proc. International Conference on Multimodal Interaction (ICMI)*, pp. 293–294 (2012).

[13] Sugiyama, T., Funakoshi, K., Nakano, M. and Komatani, K.: Estimating Response Obligation in Multi-Party Human-Robot Dialogues, *Proc. IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, Seoul, South Korea, pp. 166–172 (2015).

[14] Matsuyama, Y., Akiba, I., Fujie, S. and Kobayashi, T.: Four-participant Group Conversation: A Facilitation Robot Controlling Engagement Density as the Fourth Participant, *Computer Speech & Language*, Vol. 33, No. 1, pp. 1–24 (2015).

[15] Lala, D., Milhorat, P., Inoue, K., Zhao, T. and Kawahara, T.: Multimodal Interaction with the Autonomous Android ERICA, *Proc. International Conference on Multimodal Interaction (ICMI)*, pp. 417–418 (2016).

[16] Hirayama, T., Sumi, Y., Kawahara, T. and Matsuyama, T.: Info-concierge: Proactive multimodal interaction through mind probing, *The Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2011)* (2011).

[17] 中村和晃, 角所考, 正司哲朗, 美濃導彦, 澤木美奈子, 南泰浩, 前田英作: 擬人化エージェントとの音声対話時におけるユーザの非言語動作からの難/易及び興味/退屈の推定, 電子情報通信学会論文誌, Vol. J95-A, No. 1, pp. 85–96 (2012).

[18] Chiba, Y., Ito, M., Nose, T. and Ito, A.: User Modeling by Using Bag-of-Behaviors for Building a Dialog System Sensitive to the Interlocutor's Internal State, *Proc. Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, pp. 74–78 (2014).

[19] 冨増紗也華, 荒木雅弘: 雑談対話におけるマルチモーダル情報からの興味の有無の判定, 人工知能学会第 30 回全国大会 (2016).

[20] Higashinaka, R., Funakoshi, K., Araki, M., Tsukahara, H., Kobayashi, Y. and Mizukami, M.: Towards Taxonomy of Errors in Chat-oriented Dialogue Systems, *Proc. Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, pp. 87–95 (2015).

[21] 「パーソナルデータの利用・流通に関する研究会」報告書 (2013).
http://www.soumu.go.jp/main_content/000231357.pdf