大阪大学マルチモーダル対話コーパス Hazumi 概要（オンライン収録版）

駒谷 和範
大阪大学 産業科学研究所

岡田 将吾
北陸先端科学技術大学院大学

2023/9/7　第三者アノテーションを追加公開
2023/1/17　ダンプファイルの言語特徴量を修正
2022/6/23　初版

1　はじめに

Hazumi は大阪大学産業科学研究所で収集されたマルチモーダル対話コーパスである [1, 2]．一般から募集した人（実験参加者）と，Wizard-of-Oz (WoZ) 方式で動作するエージェントとが，いくつかの話題について目的を定めずに対話する様子が収録されている．

Hazumi には収集開始年月に対応したバージョン名を付与している．本文書は，オンラインで収録した Hazumi2010, Hazumi2012, Hazumi2105 の 3 バージョンについて説明する．対面で収録したバージョン（Hazumi1712, Hazumi1902, Hazumi1911）については，別文書[*1]や論文 [3] を参照されたい．

マルチモーダル対話システムの研究において，システムに対して一般のユーザがどのようにふるまうかを知るのは極めて重要である．しかしながら，このようなデータの収集にはまずシステムが必要であり，また個人情報を含むデータを収集することから各種の倫理的配慮が必要となるなど，研究を始める時点で既に多くの障壁がある．この障壁を緩和し，様々な分野の研究者がマルチモーダル対話システムの研究に参入できるようにしたいという願いから，本データを公開する．

これまでにも人対人の対話データはいくつか公開されているものの，人対システムの対話データは少ない．一般ユーザのふるまいは，人に対する場合とシステムに対する場合とで大きく異なり，一般ユーザがシステム開発者の期待通りにふるまうことはない．したがって，人対人ではなく，人対システムの対話に基づき，対話システムを設計することが必須である．

2020 年度と 2021 年度は，COVID-19 の影響によりオンラインでデータを収集した．収集データの概要を表 1 に示す．実験参加者が概ね 15 分間話す様子を収録し，それを交換という単位に分割して，各種のアノテーションを行った．このような大枠は対面収録である Hazumi1911 以前と同様である．また事前と事後に行った対人コミュニケーション認知に関するアンケートや，性格特性に関するアンケートは，一貫して取得している．一方で，オンラインでのデータ収集では，Kinect や生体センサによるセンシングや，振り返りアノテーションは実施していない．

なお，Hazumi という名前は，話を弾ませることができるような対話システムを作りたいという願いから名付けたものである．

表 1　収集データの概要（オンライン収録 3 バージョン）

	Hazumi2010	Hazumi2012	Hazumi2105
収集時期	2020 年 10 月開始	2020 年 12 月開始	2021 年 5 月開始
概要	人間とシステム（Wizard-of-Oz 方式）との対話 1 対話あたり 15 分〜20 分		
Wizard への指示	雑談 「実験参加者が対話を楽しむ時間が長くなるように」		
参加者	33 名（男性 17 女性 16）	63 名（男性 29 女性 34）	29 名（男性 14 女性 15）
	20 代〜60 代		
	-		実験参加 2 回目の参加者
交換数	2,798	5,334	2,235
センサ	Zoom 収録ビデオ（実験参加者とエージェントが並置）		
交換ごとのアノテーション	第三者による心象評定（7 段階；5 名[*2]） 第三者による話題継続（7 段階；5 名[*2]）		
	システム発話，システム発話の対話行為，実験参加者の発話の書き起こし		
対話全体のアンケート	実験参加者本人による事前・事後アンケート（8 段階；18 項目） Wizard による事前・事後アンケート（8 段階；3 項目） 第三者アノテータ 5 名による事前・事後アンケート（8 段階；18 項目）[*3]		
	実験参加者本人による性格特性（TIPI-J） 第三者アノテータ 5 名による性格特性（TIPI-J）[*3]		
	実験参加者本人が人間による操作に気づいたか		

2　収集環境

実験参加者と対話するシステムは，Wizard-of-Oz (WoZ) 方式で動作させた．つまり人間のオペレータが，Zoom ごしに実験参加者の様子を見ながら，専用のインタフェースを通じて，システムの応答を選択した．システムを人間が操作していたことは，実験前や実験中には実験参加者に開示せず，実験終了後に開示した．

実験参加者が対話を楽しんでいる時間が長くなるように，Wizard は発話を選択した．具体的には，実験参加者が興味がなさそうな場合に話題を変えたり，実験参加者が興味を持った様子で積極的に話している際には Wizard は聞き役に回るようにした．

実験参加者は，クラウドソーシングにより一般から報酬付きで公募された．表 1 に示されるように，Hazumi2010 では 33 名，Hazumi2012 では 63 名，Hazumi2105 では 29 名である．なお Hazumi2105 は，Hazumi2010 と Hazumi2012 で一度収録を行った実験参加者のうち，「システムが操作していたことに気づいたか」というアンケートに「気づいていなかった」と回答していた参加者に対して，再度収録を行ったものである．

各実験参加者には，事前に，研究の意義や実験参加者の権利（いつでも実験参加を撤回できることなど）を説明した同意書に同意した者のみからデータを収録した．同意書には，データ利用に関する契約が交わされることを前提として，研究者に対して研究開発目的でデータを配布できることが明記されている．顔映像などの学会発表での表示については，同意された部分のみ利用可能である．

[*1] https://www.nii.ac.jp/dsc/idr/rdata/Hazumi/documents/HazumiOverviewInPerson.pdf
[*2] 2023 年 9 月に 2 名分を追加公開
[*3] 2023 年 9 月に追加公開

YYMM	コーパスのバージョン．2010，2012，2105 のいずれか．
G	実験参加者の性別．M（男性）もしくは F（女性）．
AA	実験参加者の年代．20 から 60 まで．
NN	上記 7 文字と合わせて ID となるように付番．

図 1　ファイルの命名規則

収録されたデータの中には，住所などの個人の属性が特定される可能性がある表現を含む発話や，個人の信条嗜好を表明している発話も一部存在した．このような部分については，配布に際して，音声を無音化したり，書き起こしテキストを伏字にするなどしている．

3　データの仕様

データはその入手方法において，以下の 2 種類に大きく分けられる．

- NII IDR（国立情報学研究所 情報学研究データリポジトリ）[*4]から配布されるもの．データ利用に関する契約を交わした利用者のみに対して配布される．
  - ビデオで収録したデータ（3.1 節）
- Github からダウンロードできるもの．
  - 閲覧用 ELAN ファイル（3.2 節）
  - 実験用ダンプファイル（3.3 節）
  - アンケートデータ（3.4 節）

  URL は以下のとおりである．

  `https://github.com/ouktlab/Hazumi2010/`
  `https://github.com/ouktlab/Hazumi2012/`
  `https://github.com/ouktlab/Hazumi2105/`

ファイルは実験参加者ごとに分かれている．ファイル名には実験参加者 ID が使用されており，図 1 に示される YYMMGAANN の 9 文字（例： 2010F4001）で構成される．例えば 2010F4001 は，「Hazumi2010 の，ある 40 歳代女性のデータ」を意味する．

3.1　ビデオデータ

オンライン収録には Zoom を用いた．Zoom の画面録画機能により，実験参加者とエージェントが並ぶような画面構成で収録した．Zoom の仕様により，収録時の回線状況等によってフレームレートは異なる．

画像や映像を，論文や学会発表で使用するのは，同意がなされた項目のみ可能である．具体的には，ビデオデータに同梱される同意情報（consent.pdf）を参照のこと．

3.2　閲覧用 ELAN ファイル

アノテーションや書き起こしを全て含んだ eaf（ELAN annotation format）ファイルである．アノテーションツール ELAN[*5]上で，3.1 節で説明した実験参加者ビデオを読み込んで使用する．ELAN 5.9 で動作を確認している．

付与単位として，システムの発話と実験参加者の発話の対である交換 (exchange) を設定している．具体的

[*4] `https://www.nii.ac.jp/dsc/idr/`
[*5] `https://archive.mpi.nl/tla/elan`

には，システム発話開始時刻から，次のシステム発話開始時刻までを一交換としている．WoZ 方式での収集であるため，システムの発話開始時刻（つまり操作役の Wizard が発話開始ボタンを押した時刻）がログに記録されている．交換の終了時刻は次のシステムの発話開始時刻とした．このように機械的に交換を認定した．

この付与単位に対して，eaf ファイルの各層に，以下の各節で説明するアノテーションがなされている．

なお，一部の交換への心象アノテーションや話題継続アノテーションの値として，e や E が付与されている場合がある．これは error を意味する．これには，データ収集時のシステムの不具合などにより，発話開始ボタンを押したにも関わらずシステムが発話しなかった箇所などが含まれる．この場合ユーザは応答していないため，7 段階の数値の代わりに上記の記号を付与した．

3.2.1 実験参加者の発話の書き起こし

その交換に含まれる実験参加者の発話を，人間が聞いて書き起こしたテキストである．`user_utterance` というレイヤに記入されている．各交換内では，概ね以下の基準に基づいて書き起こしを実施した．

- 句読点は用いない
- フィラーは (F) で囲む（例：(F えーと)）
- 言い誤り・言い直しなどは (D) で囲む（例：(D ちょ）丁度いいです）
- 単位区切り記号は | (半角縦棒)
- 単位区切りの基準は 0.2 秒以上の無音（直観的には「息継ぎしたところ」）
- 明らかな文の終わりと思われるところは無音が短くても区切る

これは日本語話し言葉コーパス（CSJ）[*6]の基準のうち，基本的な部分を参考にしたものである．

3.2.2 システム発話とその対話行為

その交換のシステム発話である．これは，WoZ のインタフェース上で，Wizard が選んだものを記録したものである．`sys_utterance` というレイヤに記載されている．

さらに，システム発話の対話行為が `dialogue_act` というレイヤに付与されている．対話行為は参考文献[*7]をベースに，これを簡略化した 11 種類である（表 2）[*8]．対話行為は作業者 1 名で簡易的に付与し，別アノテータとの一致などは検証していない．

発話中に二つ以上の機能がある場合は，より後に出てくる機能を，その発話の機能とみなした．例えば，「今，「ダフィン」が流行っているようですよ．ご存知ですか？」には情報提供と Yes-No 疑問文の 2 つの機能が含まれるが，この例の発話の対話行為は，後に出てくる Yes-No 疑問文 qy とした．

3.2.3 心象アノテーション

実験参加者がシステム発話を受けてどのように感じているかを，第三者視点から，対話のビデオを前から順番に見ながら，人手で付与したものである．各交換に対して 7 段階で付与した．1 をネガティブ，7 をポジティブとした．ポジティブの例として，楽しい，話し続けたい，満足などを示し，ネガティブの例として，楽しくない，話し続けたくない，不満，困惑などを示した．5 名[*2]のアノテータにより付与し，`UI_XX` というレイヤに記入されている[*10]．`XX` は 2 文字のアノテータ ID である．

[*6] `https://pj.ninjal.ac.jp/corpus_center/csj/k-report-f/CSJ_rep.pdf`

[*7] `https://web.stanford.edu/~jurafsky/ws97/manual.august1.html`

[*8] Hazumi1902 までは表 2 中の二重線より上の 8 種類を付与していた．

[*9] 記号 fu は Follow-Up の頭文字である．例えば，システム：「どこか旅行に行きましたか？」，ユーザ：「沖縄に行きました」，システム：「沖縄ですか」の 3 発話目がこれにあたる [4]．表層的には qy にも見えるが，機能としてはユーザに Yes/No の回答を求めてはいないため，qy とはしない．

[*10] UI は User Impression の頭文字である．

表 2　システム発話の対話行為タグ

記号	内容	例文
qy	Yes-No 疑問文	スポーツはよくするんですか？
qw	Wh 疑問文	どこで食べられたんですか？
pa	肯定的回答	そうなんですか．一度乗ってみたいものですね！
na	否定的回答	すみません，知らないです．
oa	その他の回答	そうですか，いいものが見つかるといいですね．
op	開始	これから，旅行について話しましょう！
io	情報提供	人気漫画「ナルト」をテーマにした作品で，来年の夏頃に公演されるそうです．
su	提案	写真を撮る場所として，大阪の箕面の滝や和歌山の白浜などがオススメですよ！
th	感謝	きょうは実験に参加いただいてありがとうございます．
fu	了解	沖縄ですか （働きかけと応答の対の後に続く了解部分 [*9]）
no	エラー	（交換への分割が正しく行われていない箇所など）

3.2.4　話題継続アノテーション

自分がシステム役だったとした場合に，実験参加者の回答を聞いた後，次に話題を変えようと思うかどうかを人手で付与したものである．各交換に対して 7 段階で付与した．1 が「話題を変える」，7 が「この話題を続ける」を表すとした．5 名[*2]のアノテータにより付与し，`TC_XX` というレイヤに記入されている[*11]．`XX` は 2 文字のアノテータ ID である．

3.3　実験用ダンプファイル

実験用ダンプファイルは，前節までで説明した情報から特徴量を抽出し，簡便に機械学習の実験が行えるようにしたファイルである．実験参加者ごとに一つの CSV ファイルがあり，ファイル名は「実験参加者 ID.csv」である．各行が 3.2 節で述べた交換，各列がアノテーションや特徴量に対応する．

特徴量は，収録ビデオから得た，ユーザの音声，映像，言語情報（書き起こしデータ）などから抽出されたものである．特徴量セットは，これを用いて例えば，得られた実数値の特徴量を入力とし，心象・話題継続アノテーションの値（連続値とカテゴリ値）を出力とした機械学習実験が行える．

3.3.1　韻律特徴量

実験参加者の発話から openSMILE[*12]を使用して韻律特徴量を抽出した．INTERSPEECH 2009 Emotion Challenge feature set (IS09) [5] で使用された特徴量を用いた．計 384 次元の特徴量で構成されている．

3.3.2　顔表情特徴量

ビデオカメラの映像から OpenFace [6] を使用して，ランドマーク特徴量と Action Unit 特徴量の 2 種類の顔表情特徴量，計 66 次元を得た．

ランドマーク特徴量 48 次元は以下の手順で得た．まず OpenFace によりランドマーク（顔の特徴点）の 2 次元座標を求めた．具体的には，目の周りの 4 点，口の周りの 4 点，眉の周りの 4 点の合計 12 点である．これら 12 点それぞれについて 4 種類の統計量を求めた．具体的には，速度の絶対値の最大値，平均値，標準偏差と，加速度の絶対値の最大値である．速度は，フレーム t における座標データを $c(t)$，フレームインターバルを定数 I（1/30 秒）として，フレーム間速度の絶対値 $v(t) = \frac{|c(t+1)-c(t)|}{I}$ として求めた．同様にフレーム間加速度の絶対値は，$a(t) = \frac{|v(t+1)-v(t)|}{I}$ で求めた．

[*11] TC は Topic Continuance の頭文字である．
[*12] https://www.audeering.com/opensmile/

Action Unit 特徴量 18 次元として，顔表情を記述する動作単位である Action Unit (AU) の各 Unit を用いて特徴量とした．OpenFace には，18 種類の AU の有無をフレームごとに検出する学習済みモデルがある．このモデルによる検出結果を用い，交換内でそれぞれの Action Unit が検出されたフレームの割合を特徴量とした．

3.3.3 言語特徴量

ユーザの発話内容から言語特徴量を抽出した．ユーザ発話の書き起こしテキスト（3.2.1 節）から特徴量を抽出した．合計 785 次元である．各ユーザ発話に対し，1 発話内に含まれる各品詞タイプごとの単語出現回数（17 次元），および，学習済みの BERT モデルを特徴量抽出器として用いて抽出した特徴ベクトル（768 次元）を合わせて言語特徴量とした．

品詞は Stanza NLP [7] による形態素解析結果を用いて得た．品詞の種類は，ADJ: 形容詞，ADP: 等位接続詞，ADV: 副詞，AUX: 助動詞，CCONJ: 接続詞，DET: 限定詞，INTJ: 間投詞，NOUN: 名詞，NUM: 数値記号，PART: 助詞，PRON: 代名詞，PROPN: 固有名詞，PUNCT: 句読点，SCONJ: 従位接続詞，SYM: 記号，VERB: 動詞，X: その他，の計 17 個である．

BERT モデルには，京都大学で公開されている BERT 日本語 Pretrained モデル[*13] を用いた．このモデルの前処理の形態素解析では JUMAN++ が利用されている．ユーザ発話中の単語列に対する単語エンベディング列（BERT の隠れ層から得られた特徴ベクトルの列）に average pooling を行うことにより，768 次元の特徴量を得た．

3.3.4 対話的特徴量

システム発話の対話行為を 1 次元の特徴量として “Dialogue_act” 列に格納している．対話行為は，3.2.2 節の表 2 にある 11 種類である．11 次元の one-hot ベクトルに変換するなどして，対話行為の情報を機械学習に入力できる．

3.3.5 交換毎のアノテーション値

2 つのアノテーション結果が含まれている．つまり，3 名の第三者アノテータが付与したユーザ心象 (third sentiment: TS)，話題継続 (topic continuance: TC) である．TS と TC 共に 3 名によるアノテーションの値が TS1–3, TC1–3 の各列に格納されている．アノテーション内容の詳細については 3.2 節を参照のこと．

3.3.6 時刻情報

ファイル中の各列 start(exchange) end(exchange) は交換の開始・終了時間を示す．交換は，3.2 節で述べた，システム発話開始時刻から次のシステム発話開始時刻との間の区間である．これらの時間は，eaf（ELAN annotation format）ファイルに記録された交換の開始時間と終了時間に対応している．

3.4 アンケートデータ

オンライン収録 3 バージョンの全ての対話において，同じアンケートを実施した．アンケートの内容は Hazumi1911 と同様である[*14]．実施には Google フォームを利用した．なお，3.2.3 節で説明した交換単位の心象を含めたこれらの主観的アノテーション結果に対する分析は文献 [8, 9] にある．

[*13] https://nlp.ist.i.kyoto-u.ac.jp/index.php?ku_bert_japanese

[*14] 具体的な質問項目は，Hazumi1911 の Github サイトにある．
https://github.com/ouktlab/Hazumi1911/tree/master/questionnaire/

3.4.1 対話前と対話後の両方に実施したアンケート

実験参加者と Wizard の双方に対して，実験開始前と実験終了後にそれぞれ，会話者の対人コミュニケーション認知に関する測定項目 18 項目 [10] に関するアンケートを実施した．各項目は 8 段階である．

この結果は questionnaires.xlsx に保存されている．このファイル内には，実験参加者（実験前），実験参加者（実験後），Wizard（実験前），Wizard（実験後）の 4 つのタブがある．このそれぞれにおいて，実験参加者は上記の 18 項目，Wizard はこれを簡略化した 3 項目に対して，8 段階で回答した結果が記録されている．

3.4.2 対話後のみに実施したアンケート

実験参加者に対して対話後に，以下の 3 つの内容を尋ねた [11].

1 点目として，「近い将来，今回の収録で用いたような，雑談ができる AI を使ってみたいですか？」という質問に対して，7 段階の評点で回答してもらった．評点は，1 を「使ってみたい」，7 を「使いたくない」とした．さらに理由があれば自由に記述してもらった．

2 点目として「今回使用したメイちゃんは，実は AI ではなく人間が操作していました．そのことに気づいていましたか？」という質問に対して，7 段階の評点で回答してもらった．評点は，1 が「気づいていた」，7 が「気づかなかった」とした．またその補足や実験全体に対する意見や感想についても記述がある．

3 点目として，各実験参加者のパーソナリティ特性を調査した．具体的には，ビッグファイブ [12] の 5 特性を 10 項目で測定する Ten Item Personality Inventory (TIPI) の日本語版 TIPI-J [13] の質問文を利用して，実験参加者に自分自身の性格について尋ねた．

上記 3 点の結果も questionnaires.xlsx に保存されている．このファイル内の「記述式」というタブに上記の 1 点目と 2 点目が，「性格特性」というタブに上記 3 点目の結果が記録されている．

3.4.3 第三者アノテータ 5 名に対するアンケート

第三者アノテータ 5 名がビデオを見て事後的に，ラポール 18 項目と TIPI-J による性格特性をそれぞれ付与した結果を追加公開した（2023 年 9 月）．questionnaire-3rdparty-{rapport,personality}.xlsx である．5 名のアノテータによる付与結果がそれぞれのタブごとに含まれている．

3.5 前回収録時の実験参加者 ID との対応

Hazumi2105 は，このシステムとの対話が 2 回目となる実験参加者を対象として収録した．具体的には，1 回目の収録（Hazumi2010 と Hazumi2012）において，システムが人間に操作されていたことに気づいたか否かという事後アンケートに対して，気づかなかったと回答した参加者（5 以上を回答した参加者）に対して再度募集を行い，その一部が 2 回目の収録に参加した．

1 回目の収録と 2 回目の収録では別の ID が付与されているが，ID 間の対応を correspondence.csv に記載している．1 列目が 2 回目 (Hazumi2105) における ID, 2 列目が 1 回目の対話における ID である．Hazumi2105 のみに存在する．

4 問い合わせ先

ご意見・ご質問などありましたら下記までお知らせください．

〒567-0047 大阪府茨木市美穂ヶ丘 8-1
大阪大学産業科学研究所
駒谷 和範
Email komatani@sanken.osaka-u.ac.jp

謝辞

本コーパス作成の一部は，物質・デバイス領域共同研究拠点における「人・環境と物質をつなぐイノベーション創出ダイナミック・アライアンス」共同研究プログラム，JSPS 科研費 19H04171，19H05692 の助成を受けました．本コーパス作成の始点となった，人工知能学会 言語・音声理解と対話処理研究会 (SIG-SLUD)「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」のメンバー各位に感謝します．また，実験参加者の皆様やデータ収集に協力いただいた方々に感謝します．

参考文献

[1] 駒谷和範．マルチモーダル対話コーパスの設計と公開．日本音響学会誌, Vol. 78, No. 5, pp. 265–270, 2022.

[2] 駒谷和範, 岡田将吾. マルチモーダル対話コーパス Hazumi. 自然言語処理, Vol. 29, No. 4, pp. 1322–1329, 2022.

[3] Kazunori Komatani and Shogo Okada. Multimodal human-agent dialogue corpus with annotations at utterance and dialogue levels. In *Proc. Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 1–8, 2021.

[4] 荒木雅弘, 伊藤敏彦, 熊谷智子, 石崎雅人．発話単位タグ標準化案の作成．人工知能学会論文誌, Vol. 14, No. 2, pp. 251–260, 1999.

[5] Björn Schuller, Stefan Steidl, and Anton Batliner. The INTERSPEECH 2009 emotion challenge. In *Proc. Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pp. 312–315, 2009.

[6] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. Openface 2.0: Facial behavior analysis toolkit. In *Proc. International Conference on Automatic Face and Gesture Recognition (FG)*, pp. 59–66. IEEE Computer Society, 2018.

[7] Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. Stanza: A Python natural language processing toolkit for many human languages. In *Proc. Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 2020.

[8] 駒谷和範, 武田龍, 岡田将吾. マルチモーダル対話コーパスに対する主観的アノテーション結果に関する分析. 人工知能学会研究会資料, SIG-SLUD-096-40, pp. 181–186, 2022.

[9] Kazunori Komatani, Ryu Takeda, and Shogo Okada. Analyzing differences in subjective annotations by participants and third-party annotators in multimodal dialogue corpus. In *Proc. Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, 2023.

[10] 木村昌紀, 余語真夫, 大坊郁夫. 感情エピソードの会話場面における表出性ハロー効果の検討. 感情心理学研究, Vol. 12, No. 1, pp. 12–23, 2005.

[11] 駒谷和範, 岡田将吾, 堅田俊. マルチモーダル対話コーパス Hazumi 公開と生体信号を含む新規データ収集. 人工知能学会研究会資料, SIG-SLUD-C002-35, pp. 170–177, 2020.

[12] R. Lewis Goldberg. An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, pp. 1216–1229, 1990.

[13] 小塩真司, 阿部晋吾, Pino Cutrone. 日本語版 Ten Item Personality Inventory（TIPI-J）作成の試み. パーソナリティ研究, Vol. 21, No. 1, pp. 40–52, 2012.