

マルチモーダル対話コーパスHazumi

2023年12月11日（月）
@NII-IDRユーザフォーラム

大阪大学 産業科学研究所
駒谷 和範

Hazumiの特徴

- 人とシステムとの対話（雑談）
 - ユーザのふるまいは人どうしの対話とは異なる
 - システムはWizard-of-Ozで動作（人手で操作）
- ユーザのふるまいが主な収録対象
 - システムと話すユーザの表情や声の韻律，姿勢など
 - 一部バージョンでは生体信号

マルチモーダル
- 様々なアノテーションやアンケート結果
 - 参加者本人と複数（5名）の第三者アノテータが付与
 - A. 交換単位のアノテーション（全18,162交換に付与）
 - B. 対話単位のアンケート結果

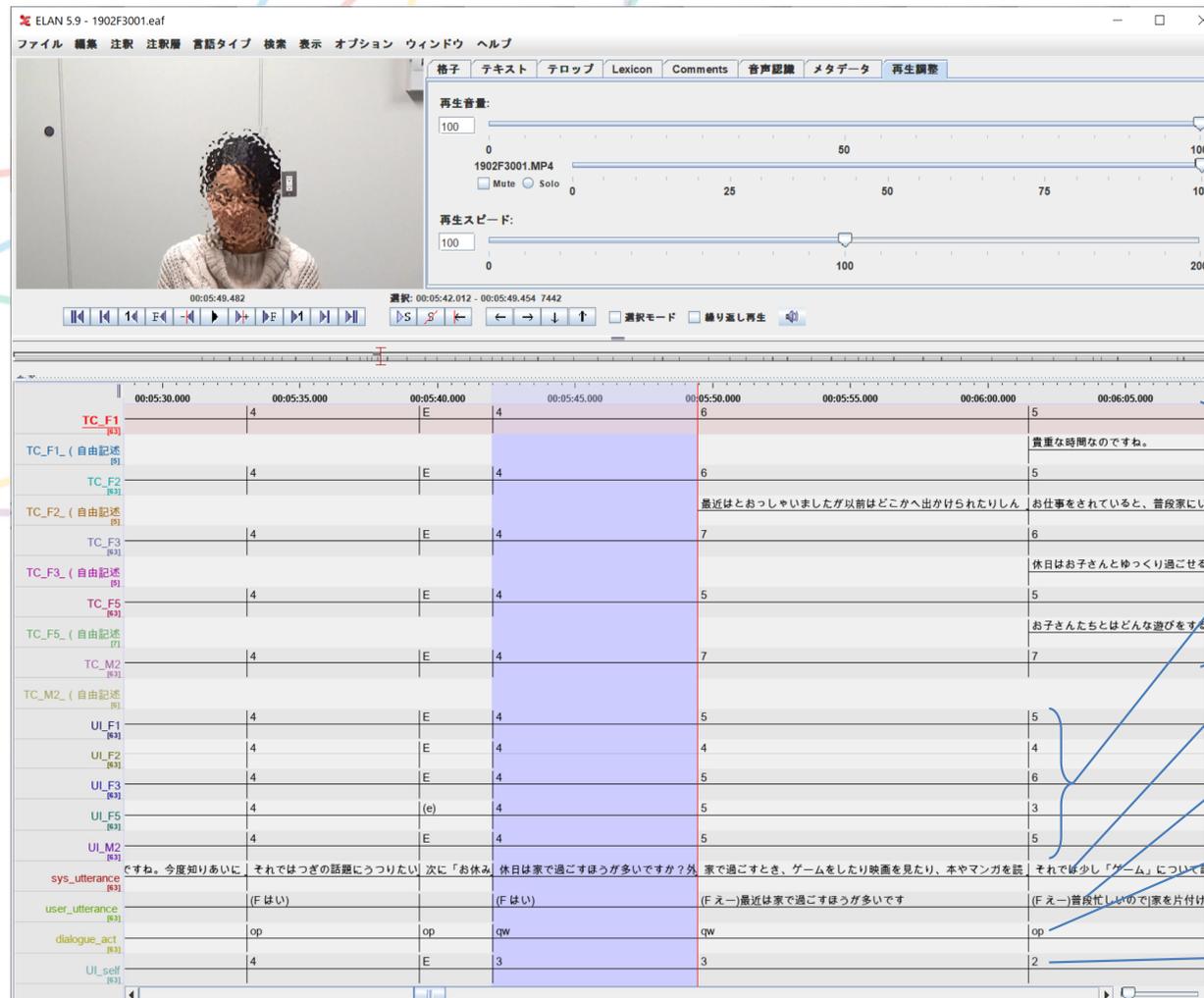
概要

- 総計214名（対話）, 18,162交換（発話対）
- 全6バージョン：バージョン番号4桁は収集開始年月

	Hazumi1712	Hazumi1902	Hazumi1911	Hazumi2010	Hazumi2012	Hazumi2105
収集開始年月	2017年12月	2019年2月	2019年11月	2020年10月	2020年12月	2021年5月
収録形態	対面収録			Zoom収録		
	ユーザとシステム（Wizard-of-Oz ; 人が裏で操作）との対話 1対話あたり15分~20分					
タスク	雑談（興味あり3話 題となし3話題）	雑談（話題：旅行, 食べ物, スポーツなど） Wizardへは「とにかく対話が盛り上がるように」と指示				
データ量 参加者	• 29名 • 2,422交換 20代~50代	• 30名 • 2,514交換 20代~70代	• 30名 • 2,859交換	• 33名 • 2,798交換	• 63名 • 5,334交換 20代~60代	• 29名（2回目） • 2,235交換
使用センサ	参加者のビデオ（顔画像+上半身） Kinect収録データ			Zoomのビデオ （参加者とエージェントが並置）		
	エージェント（システム）のビデオ					
				生体センサ		
交換単位の アノテーション	興味度（3段階； 3or6名）	参加者自身による心象評定 「振り返りアノテーション」（7段階）		心象評定（7段階；第三者5名） 話題継続（7段階；第三者5名）		
対話単位の アンケート				参加者自身によるラポール18項目（対話前, 対話後） 第三者（5名）によるラポール18項目（対話後） Wizardによるラポール3項目（抜粋）（対話前, 対話後）		
				参加者自身による性格特性（TIPI-J） 第三者（5名）による参加者の性格特性（TIPI-J） 参加者が人間による操作であることに気付いた度合		

A. 交換単位のアノテーション

全18,162交換



Time	Speaker	Segment	Label	Text
00:05:30.000	4	E	4	
00:05:40.000	4	E	4	
00:05:45.000	6		6	
00:05:55.000	5		5	貴重な時間なのでですね。
00:06:00.000	5		5	
00:06:05.000	5		5	最近はおっしゃいましたが以前はどこかへ出がけられたりしん
00:06:10.000	5		5	お仕事をされていると、普段家にい
00:06:15.000	5		5	休日はお子さんとゆっくり過ごせる
00:06:20.000	5		5	
00:06:25.000	5		5	お子さんたちとはどんな遊びをする
00:06:30.000	5		5	
00:06:35.000	5		5	
00:06:40.000	5		5	
00:06:45.000	5		5	
00:06:50.000	5		5	
00:06:55.000	5		5	
00:07:00.000	5		5	
00:07:05.000	5		5	
00:07:10.000	5		5	
00:07:15.000	5		5	
00:07:20.000	5		5	
00:07:25.000	5		5	
00:07:30.000	5		5	
00:07:35.000	5		5	
00:07:40.000	5		5	
00:07:45.000	5		5	
00:07:50.000	5		5	
00:07:55.000	5		5	
00:08:00.000	5		5	
00:08:05.000	5		5	
00:08:10.000	5		5	
00:08:15.000	5		5	
00:08:20.000	5		5	
00:08:25.000	5		5	
00:08:30.000	5		5	
00:08:35.000	5		5	
00:08:40.000	5		5	
00:08:45.000	5		5	
00:08:50.000	5		5	
00:08:55.000	5		5	
00:09:00.000	5		5	
00:09:05.000	5		5	
00:09:10.000	5		5	
00:09:15.000	5		5	
00:09:20.000	5		5	
00:09:25.000	5		5	
00:09:30.000	5		5	
00:09:35.000	5		5	
00:09:40.000	5		5	
00:09:45.000	5		5	
00:09:50.000	5		5	
00:09:55.000	5		5	
00:10:00.000	5		5	
00:10:05.000	5		5	
00:10:10.000	5		5	
00:10:15.000	5		5	
00:10:20.000	5		5	
00:10:25.000	5		5	
00:10:30.000	5		5	
00:10:35.000	5		5	
00:10:40.000	5		5	
00:10:45.000	5		5	
00:10:50.000	5		5	
00:10:55.000	5		5	
00:11:00.000	5		5	
00:11:05.000	5		5	
00:11:10.000	5		5	
00:11:15.000	5		5	
00:11:20.000	5		5	
00:11:25.000	5		5	
00:11:30.000	5		5	
00:11:35.000	5		5	
00:11:40.000	5		5	
00:11:45.000	5		5	
00:11:50.000	5		5	
00:11:55.000	5		5	
00:12:00.000	5		5	
00:12:05.000	5		5	
00:12:10.000	5		5	
00:12:15.000	5		5	
00:12:20.000	5		5	
00:12:25.000	5		5	
00:12:30.000	5		5	
00:12:35.000	5		5	
00:12:40.000	5		5	
00:12:45.000	5		5	
00:12:50.000	5		5	
00:12:55.000	5		5	
00:13:00.000	5		5	
00:13:05.000	5		5	
00:13:10.000	5		5	
00:13:15.000	5		5	
00:13:20.000	5		5	
00:13:25.000	5		5	
00:13:30.000	5		5	
00:13:35.000	5		5	
00:13:40.000	5		5	
00:13:45.000	5		5	
00:13:50.000	5		5	
00:13:55.000	5		5	
00:14:00.000	5		5	
00:14:05.000	5		5	
00:14:10.000	5		5	
00:14:15.000	5		5	
00:14:20.000	5		5	
00:14:25.000	5		5	
00:14:30.000	5		5	
00:14:35.000	5		5	
00:14:40.000	5		5	
00:14:45.000	5		5	
00:14:50.000	5		5	
00:14:55.000	5		5	
00:15:00.000	5		5	
00:15:05.000	5		5	
00:15:10.000	5		5	
00:15:15.000	5		5	
00:15:20.000	5		5	
00:15:25.000	5		5	
00:15:30.000	5		5	
00:15:35.000	5		5	
00:15:40.000	5		5	
00:15:45.000	5		5	
00:15:50.000	5		5	
00:15:55.000	5		5	
00:16:00.000	5		5	
00:16:05.000	5		5	
00:16:10.000	5		5	
00:16:15.000	5		5	
00:16:20.000	5		5	
00:16:25.000	5		5	
00:16:30.000	5		5	
00:16:35.000	5		5	
00:16:40.000	5		5	
00:16:45.000	5		5	
00:16:50.000	5		5	
00:16:55.000	5		5	
00:17:00.000	5		5	
00:17:05.000	5		5	
00:17:10.000	5		5	
00:17:15.000	5		5	
00:17:20.000	5		5	
00:17:25.000	5		5	
00:17:30.000	5		5	
00:17:35.000	5		5	
00:17:40.000	5		5	
00:17:45.000	5		5	
00:17:50.000	5		5	
00:17:55.000	5		5	
00:18:00.000	5		5	
00:18:05.000	5		5	
00:18:10.000	5		5	
00:18:15.000	5		5	
00:18:20.000	5		5	
00:18:25.000	5		5	
00:18:30.000	5		5	
00:18:35.000	5		5	
00:18:40.000	5		5	
00:18:45.000	5		5	
00:18:50.000	5		5	
00:18:55.000	5		5	
00:19:00.000	5		5	
00:19:05.000	5		5	
00:19:10.000	5		5	
00:19:15.000	5		5	
00:19:20.000	5		5	
00:19:25.000	5		5	
00:19:30.000	5		5	
00:19:35.000	5		5	
00:19:40.000	5		5	
00:19:45.000	5		5	
00:19:50.000	5		5	
00:19:55.000	5		5	
00:20:00.000	5		5	

話題継続(第三者5名;
7段階)

心象評定(第三者5名;
7段階)

システム発話

ユーザ発話(書き起こし)

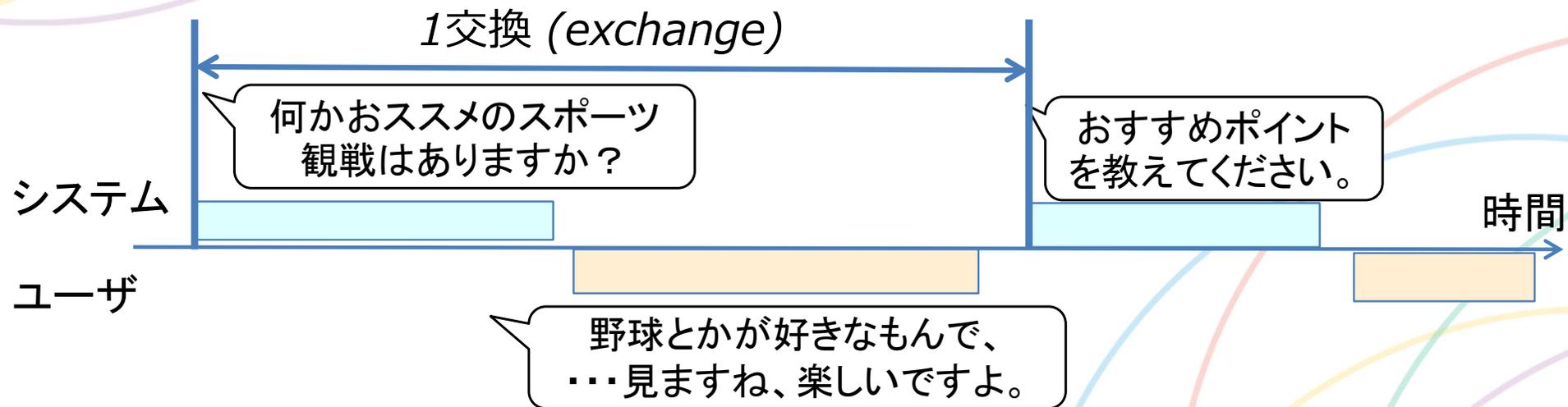
システム発話の対話行為

参加者本人による心象(振り
返りアノテーション;7段階)

- 他に実験用ダンプファイル (csvファイル) も公開
 - 各種センシング結果を含む

A. 交換単位のアノテーション

- 単位：1 交換（システムの発話－ユーザの発話）
 - 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
 - システムログから自動で認定可能



心象アノテーション（交換単位）

人と対話システムとの雑談対話データにおいて、ユーザがその直前のシステム発話を受けてどのように感じているかを、第三者の視点から付与する。

ユーザがどう感じているように思うか（7段階）

←1 2 3 4 5 6 7→
ネガティブ ニュートラル ポジティブ

ポジティブの例：楽しい、話し続けたい、満足、など

ネガティブの例：楽しくない、話し続けたくない、不満、困惑、など

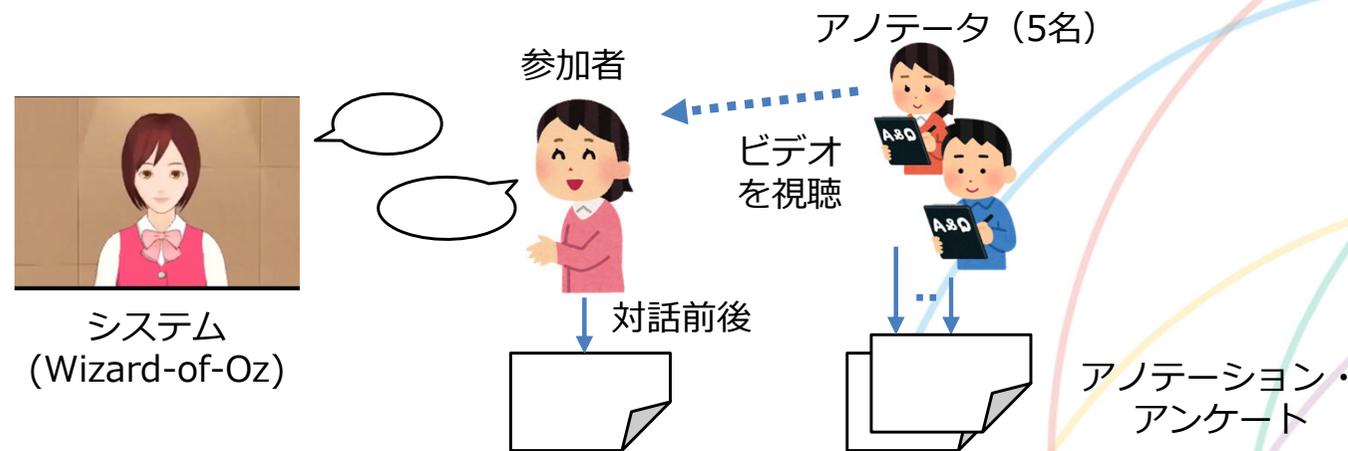
単位は1交換（「システムの発話S」-「ユーザの発話U」の対）とし、表情・音声の韻律情報・発話内容などから総合的に判断する。

ユーザが対話を楽しんでいる度合

- 第三者アノテーション：5名
- 振り返りアノテーション（参加者自身）：1名
 - 指示は「ユーザ」「第三者」を「自分」に変更

B. 対話単位のアンケート

- **ラポール18項目** [木村, 余語, 大坊 (2005)]
 - 行った対話に関する評価 (各8段階)
例: 「会話をうまく調整することができた」 など
- **性格特性**
 - TIPI-J (10項目) による Big Five (各7点×2)



参加者自身と第三者アノテータ5名が回答

ラポール18項目（対話ごと）

- 対話前／対話後に、同じ項目でアンケートを実施
- 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
 - 木村, 余語, 大坊 (2005). 感情エピソードの会話場面における表出ハロー効果の検討, 感情心理学研究, Vol. 12, pp.12-23.

実験協力者用質問紙（実験後）

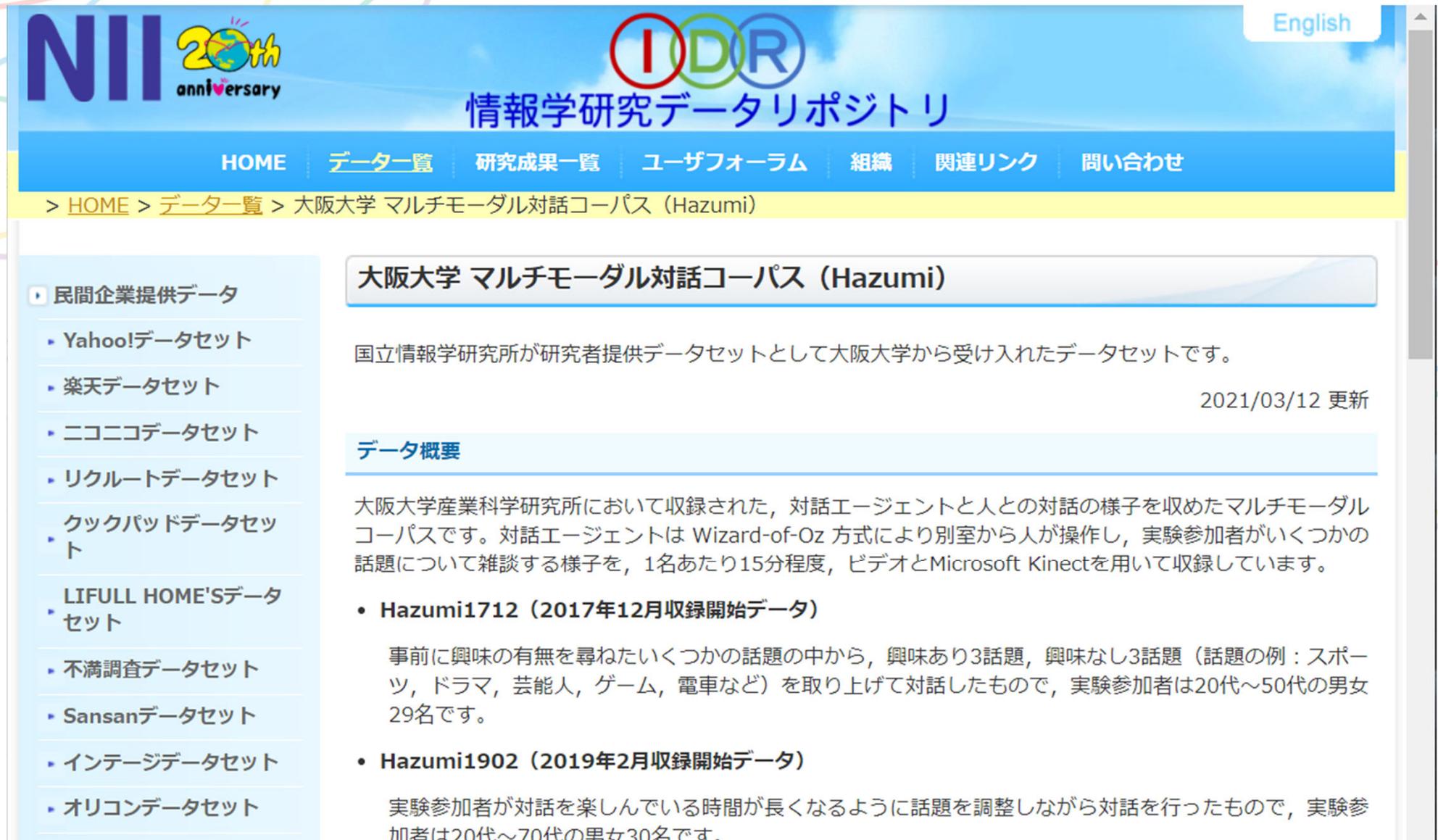
実験参加者 ID : 1902

問：システムとの対話についての質問です。以下の文章について、システムと行った対話についてはどのくらい当てはまりますか。1～8の内あてはまると思う数値に○をつけてください。

	←全くそうでない 全くその通りである→							
1. 会話をうまく調整することができた	1	2	3	4	5	6	7	8
2. 会話に退屈していた	1	2	3	4	5	6	7	8
3. 協力的に会話が進んだ	1	2	3	4	5	6	7	8
4. 調和のとれた会話だった	1	2	3	4	5	6	7	8
5. 不満足な会話であった	1	2	3	4	5	6	7	8
6. テンポの悪い会話だった	1	2	3	4	5	6	7	8

提供方法

<https://www.nii.ac.jp/dsc/idr/rdata/Hazumi/>



The screenshot shows the NII IDR website interface. At the top, there is a navigation bar with 'HOME', 'データ一覧', '研究成果一覧', 'ユーザフォーラム', '組織', '関連リンク', and '問い合わせ'. Below this is a breadcrumb trail: '> HOME > データ一覧 > 大阪大学 マルチモーダル対話コーパス (Hazumi)'. The main content area features a header for '大阪大学 マルチモーダル対話コーパス (Hazumi)' and a description: '国立情報学研究所が研究者提供データセットとして大阪大学から受け入れたデータセットです。' with a date '2021/03/12 更新'. A 'データ概要' section follows, detailing the dataset's origin and collection method. Two specific data sets are listed: 'Hazumi1712 (2017年12月収録開始データ)' and 'Hazumi1902 (2019年2月収録開始データ)', each with a brief description of the experimental setup and participant demographics.

民間企業提供データ

- Yahoo!データセット
- 楽天データセット
- ニコニコデータセット
- リクルートデータセット
- クックパッドデータセット
- LIFULL HOME'Sデータセット
- 不満調査データセット
- Sansanデータセット
- インテージデータセット
- オリコンデータセット

大阪大学 マルチモーダル対話コーパス (Hazumi)

国立情報学研究所が研究者提供データセットとして大阪大学から受け入れたデータセットです。

2021/03/12 更新

データ概要

大阪大学産業科学研究所において収録された、対話エージェントと人との対話の様子を収めたマルチモーダルコーパスです。対話エージェントは Wizard-of-Oz 方式により別室から人が操作し、実験参加者がいくつかの話題について雑談する様子を、1名あたり15分程度、ビデオとMicrosoft Kinectを用いて収録しています。

- **Hazumi1712 (2017年12月収録開始データ)**
事前に興味の有無を尋ねたいいくつかの話題の中から、興味あり3話題、興味なし3話題（話題の例：スポーツ、ドラマ、芸能人、ゲーム、電車など）を取り上げて対話したもので、実験参加者は20代～50代の男女29名です。
- **Hazumi1902 (2019年2月収録開始データ)**
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話題を調整しながら対話を行ったもので、実験参加者は20代～70代の男女30名です。

提供方法

研究開発目的に限り研究者に対して提供可能

- NIIとGithubにて公開中

- NII :

- 動画（個人情報を含むもの；要誓約書）

- <https://www.nii.ac.jp/dsc/idr/rdata/Hazumi/>

- Github :

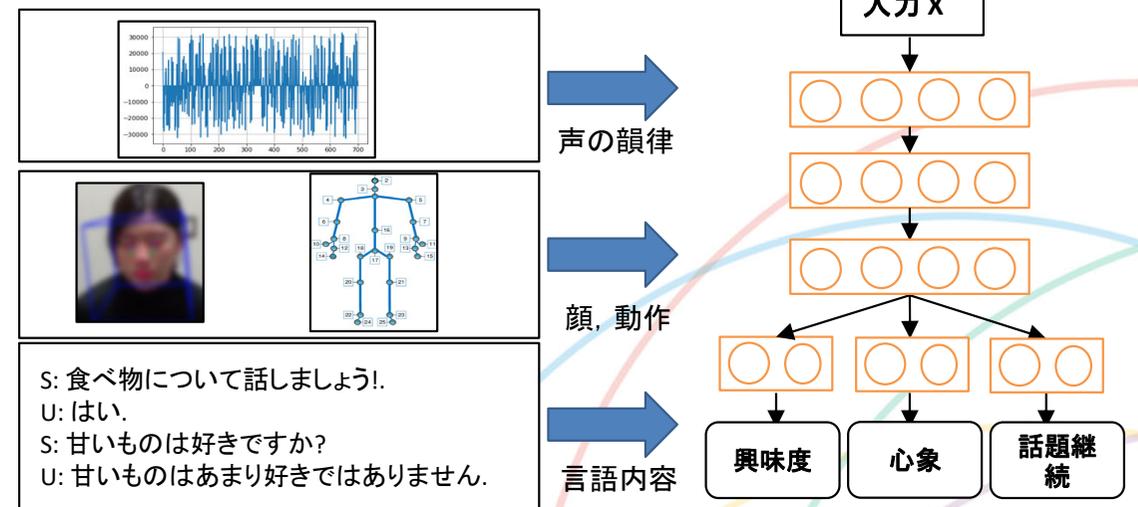
- アノテーション, 特徴量ファイルなど
（自由にダウンロード可能）

- <https://github.com/ouktlab/Hazumi>{1712,1902,...}

- 各バージョンごとにページがあります

マルチモーダル情報からの 交換単位のラベルの予測

- 複数付与されているラベルを利用
 - マルチタスク学習 [Hirano+, ICMI2019]
- 付与ラベルの曖昧性の考慮
 - 主観的に付与されたラベル
 - 弱教師あり学習 [Hirano+, ICMI2021]

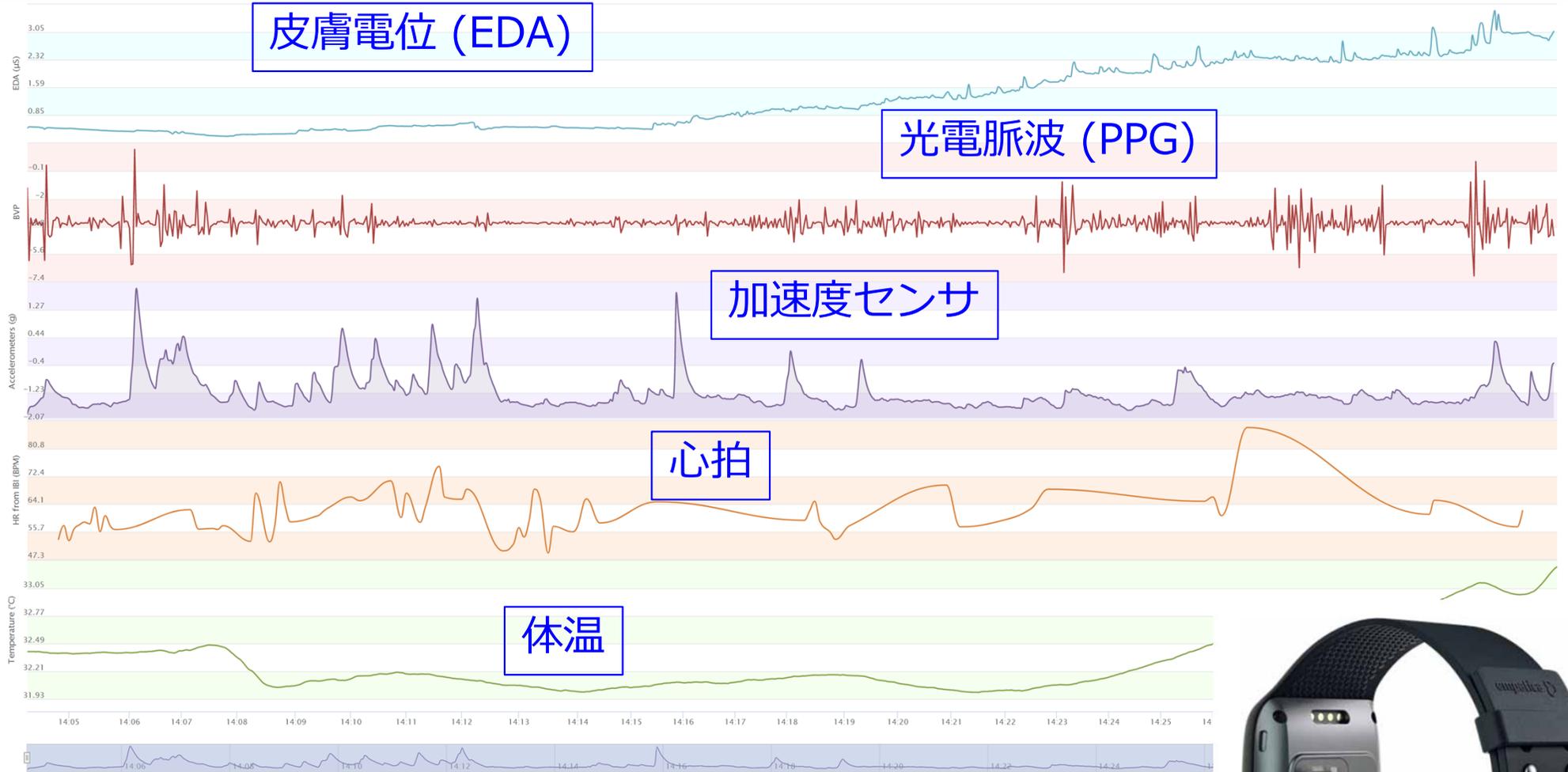


Yuki Hirano, Shogo Okada, Haruto Nishimoto, Kazunori Komatani:
Multitask Prediction of Exchange-level Annotations for Multimodal Dialogue Systems.
International Conference on Multimodal Interaction (ICMI), pp.85-94, 2019.
(Best Paper Runner-ups Award)

Yuki Hirano, Shogo Okada, Kazunori Komatani:
Recognizing Social Signals with Weakly Supervised Multitask Learning for Multimodal dialogue Systems.
International Conference on Multimodal Interaction (ICMI), pp.141-149, 2021.

生体センサ (Hazumi1911)

E4 connect Home Sessions Developer Kazunori



Empatica社
E4 wristband

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

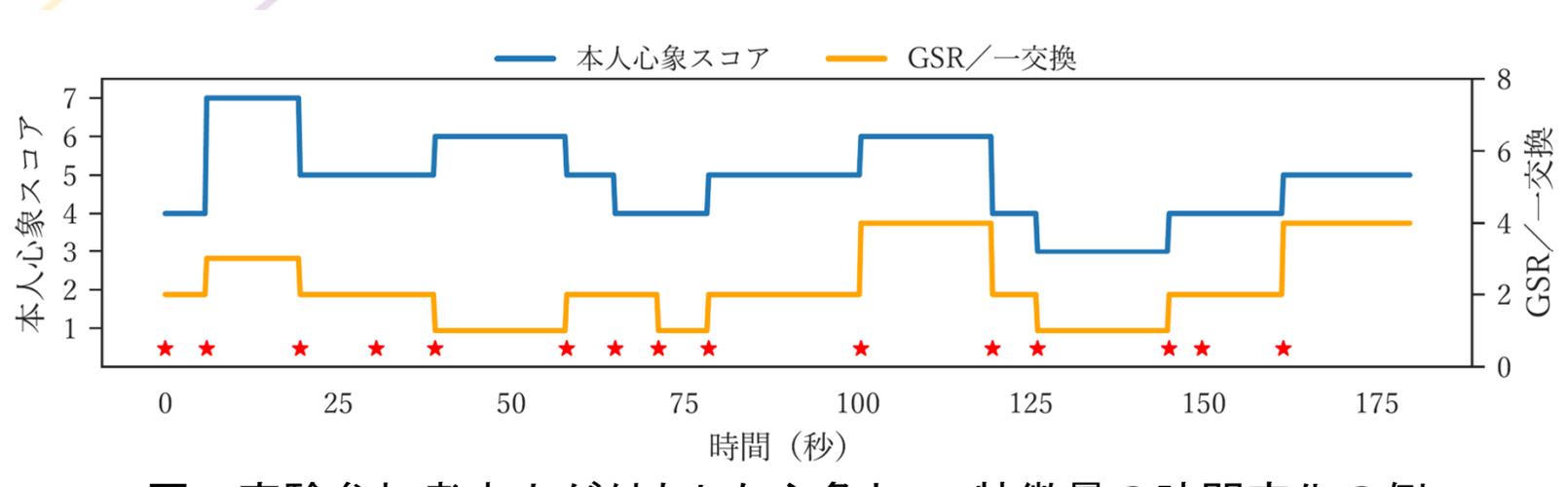
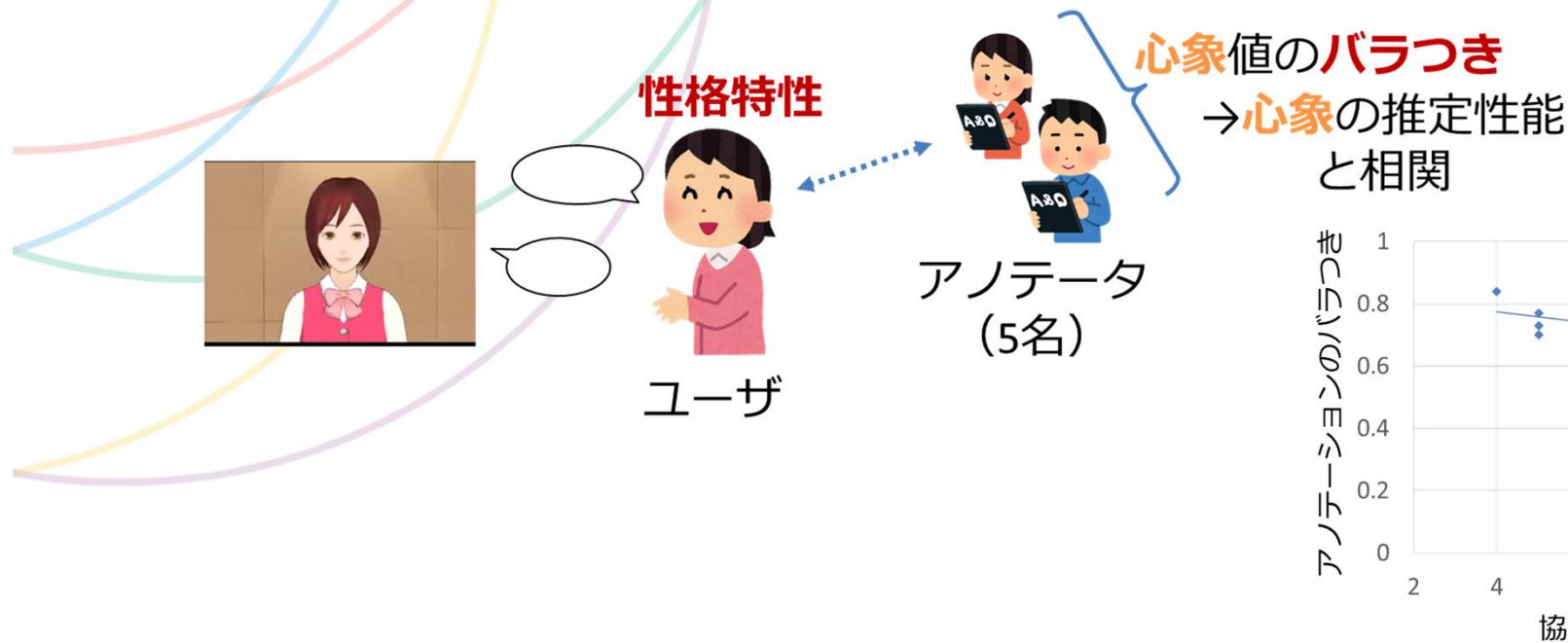


図5: 実験参加者本人が付与した心象とGSR特徴量の時間変化の例

Shun Katada, Shogo Okada, Yuki Hirano, Kazunori Komatani:
Is She Truly Enjoying the Conversation?: Analysis of Physiological Signals toward
Adaptive Dialogue Systems. Proc. ICMI, pp.315-323, 2020.

Shun Katada, Shogo Okada, Kazunori Komatani:
Effects of Physiological Signals in Different Types of Multimodal Sentiment Estimation.
IEEE Transactions on Affective Computing, Vol. 14, Issue 3, pp.2443-2457, 2023.

主観的アノテーション結果の分析



- ユーザの協調性 (本人申告) とバラつきとの間で負の相関
 - 協調性が高いユーザほど、第三者心象のバラつきは小さい
 - 他人から見てわかるように心象を表出しがち
 - 協調性の高いユーザの心象の自動推定結果は信用できる傾向にある

Kazunori Komatani, Ryu Takeda, Shogo Okada:
Analyzing Differences in Subjective Annotations by Participants and Third-party Annotators in Multimodal Dialogue Corpus. Proc. SIGDIAL, pp. 104-113, 2023.

まとめ

- マルチモーダル対話コーパスHazumi
 - 人とシステムとの対話（雑談）
 - システムはWizard-of-Oz
 - ユーザのふるまいが主な収録対象
 - 様々なアノテーションやアンケート結果
 - 交換ごと
 - **本人の**心象&第三者アノテーション
 - **生体信号**
 - 対話全体
 - 事前・事後アンケート
 - **性格特性**アンケート
 - 214名（対話）, 18,612交換
 - <https://www.nii.ac.jp/dsc/idr/rdata/Hazumi/>

