Interspeech 2022
18-22 September 2022, Incheon, Korea

Multimodal Persuasive Dialogue Corpus Using a Teleoperated Android

Seiya Kawano[1], *Muteki Arioka*[2,1], *Akishige Yuguchi*[1], *Kenta Yamamoto*[3,1], *Koji Inoue*[3], *Tatsuya Kawahara*[3,1], *Satoshi Nakamura*[2], *and Koichiro Yoshino*[1,2]

[1]Guardian Robot Project, RIKEN, Japan
[2]Graduate School of Science and Technology, Nara Institute of Science and Technology, Japan
[3]School of Informatics, Kyoto University, Japan

{seiya.kawano,akishige.yuguchi,koichiro.yoshino}@riken.jp,
{muteki.arioka,s-nakamura}@is.naist.jp, {yamamoto,inoue,kawahara}@sap.ist.i.kyoto-u.ac.jp

Abstract

We collected a corpus of persuasive dialogues containing multimodal information for building a persuasive dialogue system that encourages users to change their behaviors using multimodal information. The corpus is constructed with an android robot that was remotely controlled by the WoZ method during user interactions with the system. We transcribed the collected speech and annotated dialogue-act labels. We also extracted the facial features of the dialogue participants. Pre- and post-questionnaires identified the subjects' personality, their awareness of the target domain of persuasion, the changes in their awareness before/after the persuasion, and whether they agreed to the persuasion during the dialogues. In addition, we conducted a follow-up survey with each subject to investigate whether the persuasion actually affected their behavioral change. Moreover, we built linear classifiers that predict persuasion success to investigate effective features.

Index Terms: dialogue corpus, persuasion, multimodal

1. Introduction

Most systems with natural language interfaces, such as dialogue systems, target an actual dialogue goal (purpose), such as providing requested information [1] or summoning a target system functions [2]. Behavior changes, which is one of the most important system purposes, tries to alter a user's behavior through interactions [3]. A system that uses natural language dialogue to encourage behavior changes is a type of persuasive dialogue system [4, 5]. Such systems change user behaviors and goals through dialogues in accordance with the predefined system's objective. The common targets of persuasive dialogue systems include sales promotions [5], improving lifestyle habits [6], and addressing social issues [7].

Most existing persuasive dialogue research has focused on the linguistic aspects of dialogue, especially dialogue-acts and lexical choices [5, 8, 7, 9, 10]. In practice, however, users are influenced by the phenomena of a variety of modalities that are expressed in dialogue. For example, users are more likely to be persuaded by a person whom they trust and with whom they feel familiarity. Such trust and familiarity can be recognized from the multimodal behaviors of the users [11]. In addition, such information as the user's emotional state [12] and whether they are lying [9] can also be obtained from multimodal information to improve success of the persuasion.

In this study, we collected a multimodal corpus of persuasive dialogues with the WoZ method and a teleoperated android to begin building a multimodal persuasive dialogue system. Following existing studies [6, 7], we defined these three persuasion domains: encouraging physical exercise, reducing internet dependence, and encouraging charitable donations. We recorded the persuasion process of 60 subjects by natural dialogues between a person and a system with a humanoid android robot named ERICA [13], which served as the WoZ dialogue interface based on teleoperation [14, 15]. We recorded the subject's speech and frontal face images and extracted action units (AUs) [16] as facial features. The recorded speech was transcribed and annotated with dialogue-act tags according to the extended ISO-24617-2 dialogue-act tag standard [17]. We also annotated labels to the subjects' attributes and personality [18, 19], how they felt after being persuaded, whether they agreed with the persuasion in the dialogue, and whether actual behavioral change was induced. Questionnaires and follow-up surveys collected such information. We analyzed which factors resulted in effective persuasion using the collected data.

2. Data Collection of Persuasive Dialogues

In a persuasive dialogue system, the persuader is the system and the persuaded is the users, and the system tries to lead them to the target state based on its persuasion purpose. A dialogue ends when the user agrees with the persuasion or after a certain time (or a certain number of turns) elapses without the user's acquiescence. In our recordings, we set the minimum and maximum dialogue durations to five and eight minutes and recorded the following settings.

2.1. Persuasion Domain

We included persuasion dialogues with the goals (purposes) of encouraging exercise (exercise), reducing internet dependence (internet), and encouraging charitable donations (charity). Exercise and charity are persuasion goals that were investigated in previous studies [7, 6]. Since the existing charity system study was conducted in the United States, different results might be obtained due to different cultural attitudes to charity between the US and Japan. Therefore, we added a new persuasion goal for reducing internet dependence that might be applicable to the subjects collected in this study.

2.2. Instructions to Subjects

We collected 60 subjects (43 males and 17 females) between 18 and 38 years old for a dialogue experiment with the humanoid android ERICA. We assigned one persuasion domain for each participant. Two sessions were removed because of recording failure. We informed subjects that they are testing a new system developed with ERICA; we did not disclose that our experiment's purpose was to persuade them or that it was a WoZ dialogue. We asked our subjects to be aware of taking turns and

Copyright © 2022 ISCA

10.21437/Interspeech.2022-565

avoiding overlapping their speech with ERICA's speech. In accordance with RIKEN's ethics review, we explained that their voice and facial image features would be recorded and that personality information would be collected through questionnaires. We recruited people who were interested in exercise, the internet, or charity work and screened them in pre-surveys. They were paid after completing the follow-up survey.

2.3. Instructions to Teleoperators

A humanoid android named ERICA was remotely controlled by two teleoperators who were selected from those who had previously operated it. In addition to the instructions given to the subjects, we explained the following techniques and flow of the persuasive dialogue to the teleoperators.

The following strategies have been proven effective when conducting persuasive dialogues [7]. The teleoperators were taught in advance to effectively use them, and had trial sessions.

1. *Logical appeal*: using reason and presenting evidence to convince others.
2. *Emotion appeal*: eliciting specific emotions to influence others.
3. *Credibility appeal*: using credentials and citing organizational impacts to establish credibility and earn the persuaded's trust.
4. *Foot-in-the-door*: starting with requests for small changes to facilitate compliance followed by larger requests.
5. *Personal stories*: using narrative exemplars to illustrate someone's experiences or the beneficiaries' of positive outcomes.
6. *Actual information*: providing specific information about target tasks.
7. *Source-related inquiry*: asking whether the persuaded is aware of the problem.
8. *Task-related inquiry*: asking for the persuaded's opinion and expectations related to target tasks.
9. *Personal-related inquiry*: asking about the persuaded's previous personal experiences relevant to the task.

We removed *self-modeling* from the original strategy in our case because it does not fit the teleoperated dialogue.

The teleoperators advanced the dialogue in the following procedure:

1. The teleoperators greeted the participants.
2. The teleoperators presented the conversation topic: exercise, internet, or charity.
3. The teleoperators asked about the subject's awareness of the conversation topic (Strategy 7)
4. They proceeded with the persuasion depending on the participant's response.
5. They naturally ended the persuasion session after five minutes as much as possible.

We told the teleoperators that their payments were based on the number of dialogues they performed, and a bonus was possible based on the number of successful persuasions.

2.4. Recording Environment

The recording environment is shown in Figure 1. The voices of the subjects' and the teleoperators' were recorded in stereo using speech microphones[1]. Face images were captured from the front with a webcam[2], and the action units (AUs) were automatically annotated. ERICA changed its facial expressions during the dialogue by adjusting the valence-arousal dual axis based on the facial expressions of the teleoperators [20]. A trained annotator transcribed and annotated the recorded speech with dialogue-act tags based on the extended ISO-24617-2 dialogue-act tags [17]. Annotation detail is described in Section 3.

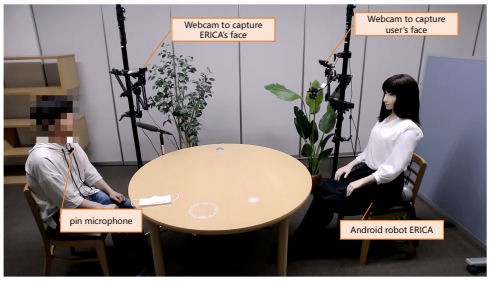

Figure 1: *Recording environment*

2.5. Questionnaires to Subjects

As a preliminary questionnaire, we surveyed subjects' personalities, decision-making tendencies, awareness of the target persuasion domains, and their current situation. This survey's items reflected those of a previous study [7]. We included questions to measure the Big Five [18, 19] to identify personalities and decision-making tendencies. As for awareness of each persuasion domain, in the exercise category, we asked about their current exercise habits and their self-awareness of such lifestyle habits. For the internet domain, we included a question on their current internet usage and their self-awareness of it. For the charity domain, we asked them about their previous charitable donations and their awareness of such participation. In accordance with previous studies, the post-questionnaires included questions on age, gender, education, partner status, and political views. In addition, we asked about the subjects' impressions of ERICA during the dialogues and how their attitudes to the target domain changed during the dialogues.

2.6. Follow-up Survey

In persuasive dialogues, it is important to know whether agreement was actually conducted after the dialogue in addition to agreement on it. We followed up a week later with the subjects who agreed with the dialogue's persuasion to see whether they had actually fulfilled such agreement. They indicated the amount of time they had spent exercising, the amount of time they had spent on the internet, and the amount of money they had donated to charity during the past week since the dialogue.

3. Corpus Annotation

We analyzed the collected corpus by discussing the success or failure of the persuasion. We discussed relations to dialogue-acts (DAs) and action units (AUs).

3.1. Persuasion Success

The persuasion's success or failure is shown in Tables 1, 2, and 3. We defined three indicators of persuasion success: change in awareness, change in behavior, and agreement through a di-

[1] SHURE, SVX188

[2] logicool, BRIO ULTRA HD PRO

Table 1: *Change in awareness*

Domain	Ratio	(number)	Changes
Exercise	95.2%	(20/21)	117.62 mins
Internet	95.0%	(19/20)	-12.80 hours
Charity	63.2%	(12/19)	-

Table 2: *Change in behavior*

Domain	Ratio	(number)	Changes
Exercise	52.4%	(11/21)	59.04 mins
Internet	50.0%	(10/20)	-0.85 hours
Charity	0.0%	(0/19)	-

alogue. Table 1 shows the changes in awareness as the ratio of subjects who answered that their interest in the target persuasion domain increased due to the dialogue and its changed value. Such changes are calculated from the post-questionnaires after the dialogue. "Changes" indicates the average questionnaire values in which the participants were asked how much time they were willing to invest. We did not ask the participants about the specific values of charity because no participants had ever donated money to charity before the dialogue. Table 2 indicates the number/percentage of subjects who answered that they actually made a behavioral change in their follow-up survey responses. Here, *changes* is the mean value of the amount of behavior change. For charity, the number of people who actually participated in charity was zero. Table 3 shows the number/percentage of subjects whoever agreed after the persuasion.

These results suggest that many people agreed to a dialogue's persuasion, regardless of any actual change in their behavior or awareness. This indicates difficulties to predict persuasion outcome only from linguistic features. To determine the effectiveness of persuasion, we must investigate whether any change in awareness or behavior was observed. Although the participants declared a change in their awareness, only half actually changed their behavior. Even if people are convinced to change their behavior through a dialogue, hurdles remain that must be removed before people take action. People must be persuaded at a time when they can immediately change their behavior to effectively promote behavioral changes.

3.2. Dialogue-Act

Dialogue-act labels were annotated to capture the trends in the actions taken by each dialogue participant. One trained annotator used the ISO-24617-2 Dialogue-Act [17] to label the

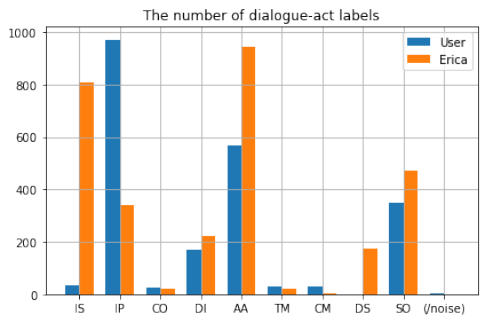

Figure 2: *Dialogue-act label distribution*

Table 3: *Agreement in dialogue*

Domain	Ratio	(number)
Exercise	90.5%	(19/21)
Internet	85.0%	(17/20)
Charity	94.7%	(18/19)

dialogue-act. Their label distribution is shown in Figure 2. Many questions were asked by ERICA (information seeking: IS), and users tended to answer them (information providing: IP). Although suggestions for mutual action (directive: DI) were observed with respect to the content of the persuasion, very little was observed with respect to commissive actions (commissive: CO). Some other dimensions of specific functions (auto/allo-feedback: AA and social obligation: SO, turn management: TM, own/partner communication management: CM, discourse structure management: DM) were also observed, which facilitated first meeting dialogues.

3.3. Action Unit

The facial action coding system (FACS) is a comprehensive, anatomically based scheme for describing all visually discernible facial movements [16]. It breaks down facial expressions into individual components of muscle movements, called action units (AUs). AUs have been used in many empirical studies that have shown its effectiveness [21]. They have also been applied extensively in face recognition and facial expression generation for conversational agents, including robots, because combinations of AUs enable an objective and concise description of a wide variety of facial expressions. We extracted the facial regions from each frame of the recorded video of a subject's behavior and automatically annotated them with AUs to identify the changes in their facial expressions at a fine-grained level. We used the open-source library PyFeat[3] for the face extraction and AU predictions. However, we only used 20 AUs as prediction targets due to the limited available automatic prediction models for a complete set of AUs[4].

4. Corpus Analysis

We examined what factors are effective in persuasion based on both pre- and post-evaluations of persuasion obtained from user questionnaires and trends in the verbal and nonverbal behaviors of users and ERICA during the persuasion. We trained linear support vector machines (SVMs) to binarily predict the success/failure of persuasion using the following five factors observed during the persuasion tasks as features.

- Per (Personality): Results of answers to the Big-Five personality traits of the 25-question and the 4-question on decision-making tendency [7, 19].
- Imp (Impression): Results of answers to 8-questions about the impressions of the robot held by the subjects through the persuasion.
- Emo (Emotion): Facial expression parameters of ERICA, manually set by the operator during the dialogue. We discretized the parameters into 9 emotion labels and counted their frequency.

[3] `https://py-feat.org/`

[4] If you need data, please contact the first or the last author. We will distribute the timestamps and the AU predictions for each frame in the videos instead of the raw persuasion videos to address any privacy issues in accordance with the ethical review results of our organization.

Table 4: *Features with high weights (top-10) of persuasion success predictors*

Model	Important Features
Personality	I devote a lot of time to others. (-1.10), I don't like being the center of attention. (0.79), I am relaxed most of the time. (-0.64), I don't mind being the center of attention. (0.63), I am interested in others. (0.63), When making decisions, I take time to consider the pros/cons and risks/benefits of the situation. (-0.63), I don't talk to people much. (-0.53) I am often grumpy. (-0.53), I tend to empathize with other people's feelings. (0.53) When I make decisions, I focus more on feelings than on analysis. (0.53)
Impression	She tried to pressure me. (-0.69), She conveyed an intention to change my behavior. (0.48). She was telling you something important. (0.38), She tried to manipulate you. (0.34), I was persuasive in her remarks. (1.14), She was dishonest. (0.14), The story was generally agreeable (0.13), The negative aspects were greatly exaggerated. (0.12)
Emo	system's neutral (0.37), user's neutral (0.37), system's relaxed (0.24), user's relaxed (0.24), system's anger (0.24), user's anger (0.24), system's bored (-0.20), user's bored (-0.20), system's sad (-0.20), user's sad (-0.20)
AU	system's avg-AU17 (-1.28), system's avg-AU06 (0.82), user's std-AU14 (0.80), system's std-AU07 (-0.76), user's std-AU12 (0.74), user's std-AU04 (0.73), system's avg-AU07 (0.73), system's avg-AU05 (0.71), system's std-AU26 (0.71), user's avg-AU25 (-0.67)
Dialogue-Act	system's information seeking (S-IS) (-1.36), user's information seeking (U-IS) (-0.98), system's social management (S-SO) (0.71), user's information providing (U-IP) (0.66), system's commissive (S-CO) (0.64), system's information providing (S-IP) (-0.44), user's commissive (U-CO) (-0.39), system's directive (S-DI) (0.30), system's auto/allo-feedback (S-AA) (0.27), user's auto/allo-feedback (S-AA) (0.21)

Table 5: *Persuasion success predictions results by SVM using RFE. Parentheses show results without using RFE.*

Model	Closed-Acc.	Eval-Acc.	Weighted-F1
Per	77.9 (82.8)	67.1 (49.1)	67.0 (48.9)
Imp	67.2 (72.4)	64.2 (54.2)	64.1 (53.9)
Emo	58.6 (62.1)	58.2 (37.3)	50.2 (36.6)
AU	78.1 (91.4)	66.7 (46.8)	66.6 (46.7)
DA	82.3 (82.8)	66.7 (60.8)	66.7 (60.7)
AU+DA	86.0 (94.8)	68.9 (45.4)	68.9 (45.1)
ALL	87.2 (100.0)	74.7 (48.4)	74.5 (48.2)

- AU (Action unit): Average and variance of the predicted AUs of the subject's facial expressions in each frame.
- DA (Dialogue-act): Frequency of dialogue-act labels in the persuasion. Since many labels did not often appear in the dialogue, we used a set of dialogue-acts at the functional level (general purpose functions: four dimensions, dimension specific functions; five dimensions).

We extracted effective features from the trained classifier to investigate feature effects. We built a set of features for Emo, AU, and DA by dividing them into two cases according to the speech intervals of users and ERICA. In addition, we applied standardization to the dimensions of feature vectors. We determined the success/failure of the persuasion based on whether the subject "actually" made a behavioral change after the persuasion ended. However, since no subject made a behavioral change concerning charity, we trained the prediction model as a task to predict those who changed their awareness.

Table 5 shows the performance of each set of features when predicting the success/failure of persuasion using closed test (Closed-Acc.) and 6-fold cross validation (Eval-Acc. and Weighted-F1). When using a single set of features, the accuracy of some prediction models did not outperform the chance rate (53.4%) that is the case of always predicting persuasion success. Moreover, the results in the closed test may indicate overfitting. Thus, we performed recursive feature elimination (RFE). The results showed that the models with feature selection improved performance. We achieved comparable performance with verbal and non-verbal features (dialogue-act labels and AUs), and got additional improvement by using both. Here, AU can be extracted automatically; thus, the fact that the result of AU can improve the result of AU+DA is also important. We achieved the best scores using all the types of feature sets (ALL). This indicates that in terms of persuasion success/failure, we must consider a combination of such factors as the attributes of dialogue participants and verbal and nonverbal information.

Table 4 shows the features with high weights (top-10) of a persuasion success predictor, trained by the overall datasets for each set of features. Here, all features were listed if the dimension of the feature vector was smaller than 10. First, we showed that users must feel that the agent is telling them something important, is trying to change their behavior, and is not trying to pressure them. We also confirmed that the following other characteristics are also related to the likelihood of persuasion: *preferring to be involved with others*, and *relying on intuition without much deliberation*. At the action level, the results suggest the importance of dialogue-acts related to self and other evaluations from both the system and the user and using dialogue-acts related to the information-providing function by the user. The role of utterances in commisive function is also important due to the nature of the task of persuasion. These results indicate that for the future construction of persuasive dialogue system, it is important to focus on making the system side's statements persuasive and elicit from users the required cues for more progress toward achieving persuasion. Such strategies as logical persuasion, appealing to credibility, and providing specific information to improve sincerity, trustworthiness, etc. contribute to a successful persuasion. Analyzing user personalities immediately before or early in the dialogue and adapting strategies accordingly may also be promising. Regarding nonverbal behavior, it is important to consider both the facial expression of the participant while speaking and the facial expression of the participant while listening to the other.

5. Conclusion

We collected a corpus of persuasive dialogues containing multimodal information for building a persuasive dialogue system that encourages users to change their behaviors using multimodal information[5]. After analyzing our corpus, we gained insight into how to judge the success or failure of persuasion and identified the factors that are important for successful persuasion. Our future work will continue to analyze multimodal information and utilize the recorded data to construct a persuasive dialogue system that uses multimodal information.

6. Acknowledgement

A part of this work is supported by JSPS KAKENHI 20H05567, 22H04873 and 22K17958.

[5]This work was reviewed by RIKEN Research Ethics Committee III and has been approved.

7. References

[1] P. Budzianowski, T.-H. Wen, B.-H. Tseng, I. Casanueva, S. Ultes, O. Ramadan, and M. Gasic, "Multiwoz-a large-scale multi-domain wizard-of-oz dataset for task-oriented dialogue modelling," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP2018)*, 2018, pp. 5016–5026.

[2] S. Tanaka, K. Yoshino, K. Sudoh, and S. Nakamura, "ARTA: Collection and classification of ambiguous requests and thoughtful actions," in *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL2021)*, 2021, pp. 77–88.

[3] B. J. Fogg, "Captology: the study of computers as persuasive technologies," in *CHI 98 Conference Summary on Human Factors in Computing Systems*, 1998, p. 385.

[4] P. Andrews, S. Manandhar, and M. De Boni, "Argumentative human computer dialogue for automated persuasion," in *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue (SIGDIAL2008)*, 2008, pp. 138–147.

[5] T. Hiraoka, G. Neubig, S. Sakti, T. Toda, and S. Nakamura, "Learning cooperative persuasive dialogue policies using framing," *Speech Communication*, vol. 84, pp. 83–96, 2016.

[6] S. Asai, K. Yoshino, S. Shinagawa, S. Sakti, and S. Nakamura, "Eliciting cooperative persuasive dialogue by multimodal emotional robot," in *Proceedings of the 12th International Workshop on Spoken Dialog System Technology (IWSDS2021)*, 2021.

[7] X. Wang, W. Shi, R. Kim, Y. Oh, S. Yang, J. Zhang, and Z. Yu, "Persuasion for good: Towards a personalized persuasive dialogue system for social good," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL2019)*, 2019, pp. 5635–5649.

[8] H. He, D. Chen, A. Balakrishnan, and P. Liang, "Decoupling strategy and generation in negotiation dialogues," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP2018)*, 2018, pp. 2333–2343.

[9] T. T. Nguyen, K. Yoshino, S. Sakti, S. Nakamura *et al.*, "Dialog management of healthcare consulting system by utilizing deceptive information," *Transactions of the Japanese Society for Artificial Intelligence*, vol. 35, no. 1, pp. DSI–C_1, 2020.

[10] W. Shi, X. Wang, Y. J. Oh, J. Zhang, S. Sahay, and Z. Yu, "Effects of persuasive dialogues: testing bot identities and inquiry strategies," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI2020)*, 2020, pp. 1–13.

[11] Y. Yamazaki, Y. Chiba, T. Nose, and A. Ito, "Construction and analysis of a multimodal chat-talk corpus for dialog systems considering interpersonal closeness," in *Proceedings of the 12th Language Resources and Evaluation Conference (LREC2020)*, 2020, pp. 443–448.

[12] N. Lubis, S. Sakti, K. Yoshino, and S. Nakamura, "Eliciting positive emotion through affect-sensitive dialogue response generation: A neural network approach," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI2018)*, vol. 32, no. 1, 2018.

[13] K. Inoue, P. Milhorat, D. Lala, T. Zhao, and T. Kawahara, "Talking with erica, an autonomous android," in *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 2016, pp. 212–215.

[14] N. M. Fraser and G. N. Gilbert, "Simulating speech systems," *Computer Speech & Language*, vol. 5, no. 1, pp. 81–99, 1991.

[15] S. Nishio, H. Ishiguro, and N. Hagita, "Geminoid: Teleoperated android of an existing person," *Humanoid robots: New developments*, vol. 14, pp. 343–352, 2007.

[16] P. Ekman, W. V. Friesen, and J. C. Hager, *Facial Action Coding System*, 2nd ed. Salt Lake City, USA: Research Nexus, 2002.

[17] K. Yoshino, H. Tanaka, K. Sugiyama, M. Kondo, and S. Nakamura, "Japanese dialogue corpus of information navigation and attentive listening annotated with extended ISO-24617-2 dialogue act tags," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association (ELRA), May 2018. [Online]. Available: https://aclanthology.org/L18-1462

[18] O. P. John, E. M. Donahue, and R. L. Kentle, "Big five inventory," *Journal of Personality and Social Psychology*, 1991.

[19] L. R. Goldberg, "The development of markers for the big-five factor structure." *Psychological assessment*, vol. 4, no. 1, p. 26, 1992.

[20] A. D., P. F.J., and V. J., "Mood and its mapping onto facial expressions." *Articulated Motion and Deformable Objects. AMDO 2014. Lecture Notes in Computer Science*, vol. 8563, 2014.

[21] B. Martinez, M. F. Valstar, B. Jiang, and M. Pantic, "Automatic analysis of facial actions: A survey," *IEEE transactions on affective computing*, vol. 10, no. 3, pp. 325–347, 2017.