Personality Classification from Robot-mediated Communication Cues*

Oya Celiktutan[1], Paul Bremner[2] and Hatice Gunes[1]

I. INTRODUCTION

Robotic telepresence offers a convenient substitute for face-to-face communication by enabling individuals to engage in communication regardless of location, and in various scenarios such as remote education, business meetings and elderly care. In robot-mediated communication, a teleoperator conveys nonverbal communication cues such as head movements, hand gestures, body posture along with audio cues through a robot. Correctly interpreting these nonverbal cues plays an important role in forming impressions, understanding social behaviours and achieving an effective communication. However, it is difficult to form a holistic understanding of how these nonverbal communication cues are interpreted by the interlocutor along with the robot's appearance, and how the perceptions regarding the teleoperator change as compared to communication in-person.

This paper focuses on automatic personality classification from nonverbal communication cues in a telepresence context and provides a comparison with respect to in-person communication condition as illustrated in Fig. 1. We extract a rich set of features and learn the relationship between the extracted features and the personality assessments by training automatic classifiers. Our results show that personality classification from robot-mediated communication cues works better than from audio-only cues except for *agreeableness*. Facial activity and head pose together with audio and arm gestures play an important role in conveying *extroversion* and *agreeableness*.

II. AUTOMATIC PERSONALITY CLASSIFICATION

This work is the continuation of our study described in [1] where we investigated human personality perception of robot avatar operators. We used a robotic telepresence platform where human gestures were replicated on a humanoid robot using a motion capture system. In this paper, we use this dataset and perform automatic personality classification. The personality classification provides insight into (i) how nonverbal communication cues such as speaking style, hand and arm movements can be utilised for robotic telepresence; (ii) whether the personality of a teleoperator is reliably inferred when a limited set of nonverbal communication cues are available to the interlocutor, and (iii) whether the robot-mediated communication provides better information than the audio-only communication for the classification of personality.

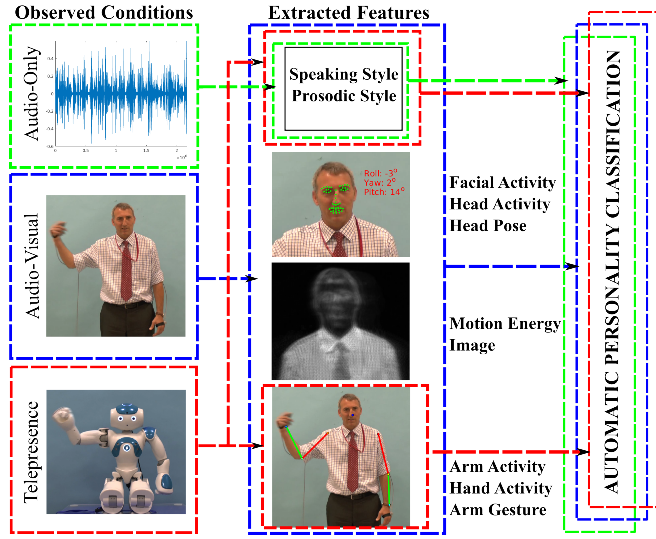

Fig. 1. Automatic personality classification under three communication conditions: Audio-Only (green dashed lines), Audio-Visual (blue dashed lines) and Telepresence (red dashed lines).

A. Data and Labels

A total of 20 participants were asked to perform two different tasks (i.e., story and hobby) and were recorded by RGB video camera [1]. From these clips, we created three communication conditions: (i) audio-only (AO - participants were not visible); (ii) audio-visual (AV); and (iii) telepresense (TP). This resulted in a total of 120 clips with mean duration = 50 s and standard deviation = 20 s. For each clip, we collected independent assessments of personality along the widely known Big Five personality traits [2] from external observers (i.e., judges). Ground-truth (i.e., personality labels) was generated by taking the average of reliable judges' assessments per clip (see [1] for details).

B. Feature Extraction

We automatically extracted a rich set of features to encode various nonverbal communication cues namely, audio, face/head and body features as illustrated in Fig. 1.

Audio Features. We extracted two sets of vocal features using a speech feature extraction tool to model *speaking activity* and *prosodic style*, similarly to our previous work [3].

Face/Head Features. We tracked landmark points and estimated head pose as in [3]. In addition to *head pose*, we considered *head activity* and *facial activity* features to capture facial/head cues.

Body Features. As appearance features, we computed *Motion Energy Images (MEIs)*, where the intensity values

*This research work is funded by the EPSRC under its IDEAS Factory Sandpits call on Digital Personhood (Grant Ref.: EP/L00416X/1).

[1]O. Celiktutan and H. Gunes are with the Computer Laboratory, University of Cambridge, UK. {Oya.Celiktutan.Dikici, Hatice.Gunes}@cl.cam.ac.uk

[2]P. Bremner is with the Bristol Robotics Laboratory, University of West England, UK. paul.brember@brl.ac.uk

indicate the amount of motion occurred at each pixel location over the whole clip. From the MEIs, we calculated simple statistics (i.e., mean, standard deviation) and number of nonzero pixels. As geometric features, we estimated upper body articulated pose (skeleton) using the deep convolutional neural networks-based method of [4] (see Fig. 1). From the estimated joint positions, we computed *arm activity* and *hand activity* features. We also represented each clip as a histogram of key *arm gestures* using a bags-of-words approach.

C. Classification

We formulated the classification task as a binary classification problem and divided the clips into two classes (e.g., extroverted vs. introverted, emotionally stable vs. neurotic, etc.). We used three different classification methods (i.e., Ridge regression, linear and nonlinear Support Vector Machines) with nested leave-one-subject-out cross validation strategy. Our results showed that different methods worked well in conjunction with different types of features.

III. Experimental Results

We used the abovementioned classification methods to learn the relationships between the extracted features and the personality labels for classifying the teleoperator's personality. The best classification results ($\geq 60\%$) were highlighted in bold for each feature type with respect to different communication conditions in Table I. In the audio-only condition, we obtained a modest classification performance for *agreeableness* only. In the audio-visual condition, the methods failed to successfully recognise *neuroticism*. In this condition, *extroversion* and *agreeableness* were found to be the easiest traits to recognise. Prominent features were head activity, head pose, arm activity and arm gestures.

In the telepresence condition, we only considered the features that represent the nonverbal communication cues that were reproduced on the robot (i.e., audio and lower arm movements with a static torso). Although all classification methods failed in modelling *openness*, this condition yielded a considerable performance for recognising this trait. While for *extroversion*, *conscientiousness* and *neuroticism*, body features worked better, for *agreeableness*, only speaking activity features were useful.

IV. Discussion and Conclusion

In the audio-only condition, poor classification results clearly showed that audio features alone were not sufficient for automatic personality recognition except for *agreeableness* (see Table I, audio-only). Similarly, in [3], we observed that in order to obtain a more complete assessment of personality, one needs to have in hand multiple clips of the observed person from audio and visual channels together. Audio-visual condition results also supported this finding as classifiers were more successful in modelling the relationship between audio, face/head and body features for recognising *extroversion*, *agreeableness* and *conscientiousness* (see Table I, audio-visual). Especially, arm gesture (75%), face activity (67.5%), head activity (67.5%) and speaking activity (67.5%) features worked better for recognising *extroversion*. This also has implications on how the telepresence robots should be designed in order to convey the personality of the teleoperator. To convey the teleoperator's *extroversion* and *agreeableness* traits more accurately, the robot should portray head pose or facial activity together with audio and arm gestures. On the other hand, for *neuroticism* and *openness*, classifiers were unable to model the relationship between the extracted features and the observer-assessments. This shows that observers seemed not to utilise these features for *neuroticism*.

TABLE I

Classification results (%). EXT: extroversion, AGR: agreeableness, CON: conscientiousness, NEU: neuroticism, OPE: openness, $^\diamond$: ridge, †: linear SVM, ‡: nonlinear SVM.

	Features	EXT	AGR	CON	NEU	OPE
Audio	*SpeakActivity*	52.5$^\diamond$	**62.5**†	57.5†	47.5$^\diamond$	50†
	ProsodicStyle	50$^\diamond$	**60**‡	47.5†	55‡	47.5$^\diamond$
	AllAudioFeat	50$^\diamond$	**62.5**†	52.5†	55‡	45$^\diamond$
Audio-Visual	*SpeakActivity*	**67.5**†	**60**$^\diamond$	52.5$^\diamond$	55†	45‡
	ProsodicStyle	**60**$^\diamond$	52.5$^\diamond$	**67.5**$^\diamond$	57.5‡	**60**‡
	AllAudioFeat	**60**$^\diamond$	57.5$^\diamond$	**62.5**‡	**60**†	57.5‡
	HeadActivity	**67.5**‡	**65**$^\diamond$	**62.5**‡	57.5$^\diamond$	50‡
	FaceActivity	**67.5**$^\diamond$	**60**$^\diamond$	57.5$^\diamond$	50$^\diamond$	55†
	HeadPose	**60**$^\diamond$	**70**$^\diamond$	**62.5**$^\diamond$	55$^\diamond$	47.5$^\diamond$
	AllFaceFeat	**60**$^\diamond$	**67.5**†	**67.5**†	50‡	45$^\diamond$
	MEIstats	57.5†	**67.5**$^\diamond$	**60**†	55†	50$^\diamond$
	ArmActivity	**65**$^\diamond$	**70**$^\diamond$	57.5†	55†	52.5$^\diamond$
	HandActivity	**62.5**$^\diamond$	**67.5**$^\diamond$	52.5$^\diamond$	50†	**60**†
	ArmGesture	**75**†	**65**$^\diamond$	**70**‡	55$^\diamond$	**67.5**‡
	AllBodyFeat	**62.5**†	**70**$^\diamond$	57.5$^\diamond$	57.5$^\diamond$	55‡
Telepresence	*SpeakActivity*	57.5†	**60**$^\diamond$	52.5‡	55$^\diamond$	**70**‡
	ProsodicStyle	52.5$^\diamond$	50†	50†	52.5‡	**72.5**†
	AllAudioFeat	47.5$^\diamond$	47.5†	55‡	52.5$^\diamond$	**67.5**$^\diamond$
	HandActivity	57.5$^\diamond$	45†	52.5‡	**60**$^\diamond$	**65**$^\diamond$
	ArmGesture	50$^\diamond$	52.5$^\diamond$	50$^\diamond$	**62.5**‡	**62.5**†
	AllBodyFeat	**65**$^\diamond$	45$^\diamond$	**60**$^\diamond$	55$^\diamond$	**65**†

Looking at the telepresence condition results (Table I, telepresence) we observe that the use of a robot avatar for telepresence helps to discriminate between different personality types (e.g., extroverted vs. introverted) better than audio-only mediated communication for *extroversion* (65%) and *conscientiousness* (60%) . In the telepresence condition, when the observers were exposed to the robot's appearance, we observed some interesting results. Although all methods failed to recognise *openness*, we achieved a considerable performance with audio and body features for this trait in the telepresence condition. In addition, body features worked better for recognising *neuroticism* in the telepresence condition as compared to audio-only and audio-visual conditions. These results indicate that robot appearance plays an important role in conveying the teleoperator's personality.

References

[1] P. Bremner, O. Celiktutan, and H. Gunes, "Personality perception of robot avatar tele-operators," in *ACM/IEEE Int. Conf. on Human-Robot Interaction*, 2016.

[2] A. Vinciarelli and G. Mohammadi, "A Survey of Personality Computing," *IEEE Trans. on Affective Computing*, 2014.

[3] O. Celiktutan and H. Gunes, "Automatic prediction of impressions in time and across varying context: Personality, attractiveness and likeability," *IEEE Trans. on Affective Computing*, 2016.

[4] T. Pfister, J. Charles, and A. Zisserman, "Flowing convnets for human pose estimation in videos," in *IEEE Int. Conf. on Computer Vision*, 2015.