Automatic Prediction of Perceived Traits using Visual Cues under Varied Situational Context

Jyoti Joshi
Vision and Sensing Group
University of Canberra, Australia
jyoti.joshi@canberra.edu.au

Hatice Gunes
School of Electronic Eng. & Computer Science
Queen Mary University of London, UK
h.gunes@qmul.ac.uk

Roland Goecke
Vision and Sensing Group
University of Canberra, Australia
roland.goekce@ieee.org

***Abstract*— Automatic assessment of human personality traits is a non-trivial problem, especially when perception is marked over a fairly short duration of time. In this study, thin slices of behavioral data are analyzed. Perceived physical and behavioral traits are assessed by external observers (raters). Along with the big-five personality trait model, four new traits are introduced and assessed in this work. The relationship between various traits is investigated to obtain a better understanding of observer perception and assessment. Perception change is also considered when participants interact with several virtual characters each with a distinct emotional style. Encapsulating these observations and analysis, an automated system is proposed by firstly computing low level visual features. Using these features a separate model is trained for each trait and performance is evaluated. Further, a weighted model based on *rater credibility* is proposed to address observer biases. Experimental results indicate that a weighted model show major improvement for automatic prediction of perceived physical and behavioral traits.**

I. Introduction

Social interactions are highly influenced and colored by the perception of human physical and behavioral traits. Extensive literature in the social psychology suggests that the perception and assessment of personality traits involves spontaneous, unintentional and unaware processes [1]. Such early and unintentional assessment of traits often directs the behavior of an individual during interpersonal communication. In human computer interaction (HCI), the user's mood, attitude and personality have pivotal roles in marking the success of an interaction. Bickmore and Picard [2] argued that people tend to like computers more when computers match their own personality. In order to create such a successful match, interactive systems need to be equipped with the capability of automatically analyzing and predicting the perceived physical and behavioral traits of the user and engage with them accordingly. This concept has already been explored and exploited in various studies (e.g. [3] and [4]) where trait information is integrated to create more believable life–like virtual characters. The underlying idea is that virtual characters can be made more realistic by putting a constraint over their behavior inclination to match the given personality and the emotion model.

Modeling human behavior perception and its manifestation automatically is quite a challenging task. Recent studies in this domain such as [5], [6], [7], [8] have used the traits outlined by the Big Five (BF) [9] personality model, widely used in psychology, to assess human personality. The BF model encapsulate different human personality traits along five dimensions:(1) *Extraversion* describes how outgoing and energetic one is; (2) *Agreeableness* describes one's orientation towards other people and one's tendency for compliance; (3) *Conscientiousness* describes one's tendency to favor structure, order and a planned behavior; (4) *Neuroticism* describes one's tendency to negative emotions such as anxiety, hostility, depression or anger; and (5) *Openness* to experience describes one's tendency to changing experience, adventure, ideas etc. Although the BF model has been extensively used to assess personality, there exist various other physical and behavioral traits which are not always covered by the BF model, especially, when the interactions are fairly short. During short interactions certain physical and behavioral aspects may dominate and color the initial assessment of one's personality. In this study, we are interested in analyzing the thin slices of behavioral data from video clips which are on average 14s long. Along with the five standard dimensions of the BF model, we are interested in analyzing four extra dimensions: *Engagement*, *Facial Attractiveness*, *Vocal Attractiveness* and *Likability*. The *Engagement* dimension forms the basis for obtaining an understanding of user attentiveness during the interaction. Overall, we are interested in understanding how external observers attribute *Likability* in human-machine interaction settings. To study this in detail, two more dimensions *Facial Attractiveness* and *Vocal Attractiveness* are also added as these are reported to affect ratings due to the halo effect that assumes that physically attractive people are nicer and superior to others along many traits [10].

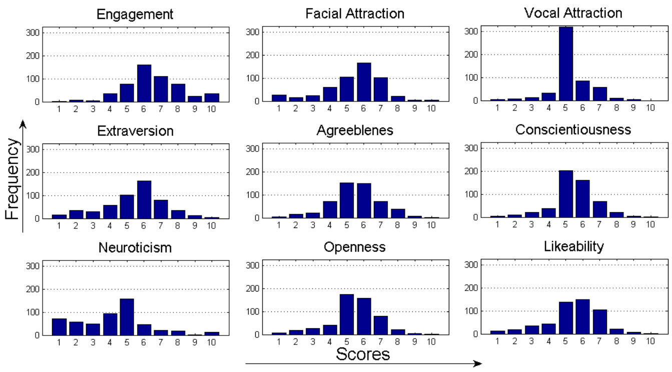


Fig. 1. Distribution of scores for 9 different dimensions.

The aim is to model the perception of traits marked by external observers of an interaction taking place between a human participant and several virtual characters each with a distinct emotional style. The first contribution of our work is investigating how various trait dimensions are inter-related.

These relationships, once understood, can be leveraged to learn and train better models towards building an automatic framework. Understanding the effect of mode of interaction is also needed in order to create a robust and efficient system. There are various works focusing either on audio [6] or visual information [11], and some on multiple information channels [8]. Accordingly, the second contribution of our work is investigating the differences between the perceived traits during audio-visual and visual only observations. A deviation in the perception is also analyzed when there is a change of situational context. Humans exhibit different social attributes while interacting with different kinds of people. Therefore, it is indeed of interest to explore the change in the perception marked by an external observer when the same individual interacts with different virtual characters manifesting varied social attributes. The third contribution of our work is a framework encapsulating a weighted model to account for the credibility of a rater for automatic prediction of perceived traits using low-level visual features.

II. Related Work

Some of the early works such as Argamon *et al.* [12] and Oberlander *et al.* [13] used textual lexical content to distinguish between personality traits of authors. Argamon *et al.* differentiated between high and low levels of Neuroticism and Extraversion using linear Support Vector Machine (SVM). Oberlander *et al.* used Naive Bayes classifier and SVM to perform binary and multiple classification using n-gram features to recognize one of the four traits of the BF model. In one of the seminal works towards automatic personality recognition in audio-visual mode, Pianesi *et al.* [8] used acoustic and visual features to predict Extraversion and Locus of Control in specific aspects of social interaction.

Vinciarelli *el al.* [14] explored the influence of non-verbal behavioral cues on social perception during zero-acquaintances scenario. They also examined the social attractiveness of the unacquainted people and its correlation with social perception and found that some of the behavioral cues such as laughter and back channel significantly influence the perception of social attractiveness. In the current work, using the findings in [14] as basis, we are interested to see how external observers attribute *Likability* in human-machine interaction settings. To study this in detail, three more dimensions *viz. Engagement, Facial Attractiveness and Vocal Attractiveness* are introduced and assessed along with the five dimensions of the BF model.

In a recent work, Batrinca *et al.* [15] proposed automatic recognition of traits of the BF model where the data was collected via a human-machine interaction task. The corpus was based on a Map task [15] where the user interacted with machine in four different levels of collaborative settings. The hypothesis of their work was that collaborative behaviors elicit the manifestation of traits related to sociability and positive outcomes (*e.g.* Agreeableness and Extraversion), while non-collaborative behaviors may trigger anxious reactions and the manifestation of related traits (*e.g.* Neuroticism) in the users. In their work, it was found that there is not a strong relationship between the differences in interactional context and different personality traits exhibited by the user. Our scenario of contextual variation differs from [15] in a way that rather than being in different collaborative settings with the computer, humans are interacting with different virtual characters which exhibit a specific kind of emotional and social attribute. Motivated by their approach, we hypothesize that people react differently and manifest different social traits when encountering a certain type of behavior in an interaction. Therefore we are interested in studying the differences in situational context affecting the trait perceptions and ratings in our data.

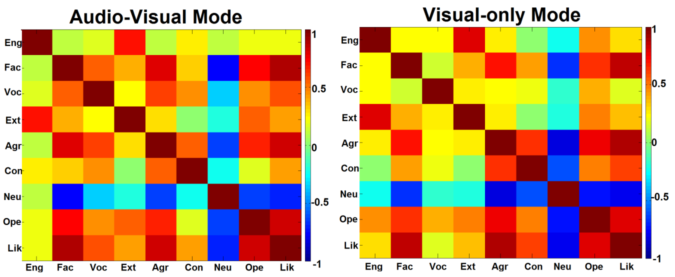
In an interesting work, Biel *et al.* [11] used facial expression analysis on subjects in online social videos and conversational vlogs to predict the personality impressions listed in the BF model. They argued that facial expressions provide information other than affective states, influencing interpersonal impressions such as personality judgements. This holds true especially in the vlogging scenario where mostly head and shoulders are visible, and the face covers the largest area on screen. A standard facial expression recognition system was used and an association of facial expressions with predicted personality traits was examined. Similar to their work, the data used in this paper also contains the upper body of the participants. Inspired by [11], features are computed only over the facial region and are further used for training a model for each personality trait.

Mohammadi *et al.* in [7] used prosodic features for personality trait prediction attributed by human listeners to unknown speakers. This study was carried out on a relatively larger data set for automatic personality prediction with 640 speech clips and 322 individuals in them. There were 11 different raters who marked the scores of perceived personality traits. One of the concerns listed by the authors was that the low agreement of the raters on the personality traits leads to erroneous results and low performance of their automatic prediction framework. As argued in the literature [16], when perceived traits are studied, there is no "*right*" or "*wrong*" perception, and disagreement or a low level agreement on a perceived attribute is rather expected. Our work addresses the low disagreement issue described in [7] by leveraging the consistency of rater assessments. Accordingly, one of the main contributions of our work is a weighted model for automatic personality prediction based on rater credibility.

III. Data and Annotation

The data set in this study is extracted from the SEMAINE corpus [17], which is freely available for scientific research purposes. The SEMAINE corpus has been recorded to study the behavioral changes and different affect manifestations by a user interacting with four "*virtual characters*", each with a distinct emotional style, and a conversational goal of shifting the user towards that state. These four characters are Prudence, even-tempered and sensible; Poppy, happy and outgoing; Spike, angry and confrontational; and Obadiah, sad and depressive. Interaction with these four different characters is termed as *varied situational context* in this work. In other words, the situational context changes along with each character displaying distinct emotional attributes. Different emotional attributes also correspond to different social traits. The goal is to elicit different types of user emotional and social behavior while user interacts with these virtual characters.

There are 44 clips in total extracted from the SEMAINE corpus. These clips consist of audio-visual recordings of 11



(a) Audio-visual (b) Visual-only

Fig. 2. Correlation between various dimensions, from top-bottom and left-right: Engagement (Eng), Facial Attractiveness (Fac), Vocal Attractiveness (Voc), Extraversion (Ext), Agreeableness (Agr), Conscientiousness (Con), Neuroticism (Neu), Openness (Ope), Likability (Lik). As per the legend markings, highly saturated red blocks signify high positive correlation and highly saturated blue blocks represent high negative correlation.

different participants interacting in four different situational contexts. In order to analyze the effect of *visual-only* behavioral cues on the perception of traits, these 44 clips are first assessed by 6 raters along the five dimensions of the BF model and the four additional dimensions employed in this work. Furthermore, to analyze the effect of *audio-visual* behavioral cues on the perception of traits, the same 44 clips are rated by the same 6 raters together with the audio channel. The dimensions were scored on a Likert scale with ten possible values, from "*strongly disagree*" to "*strongly agree*", mapped onto the range from [1,10]. All 6 raters were also shown one preselected audio-visual and one preselected visual-only clip twice, but in random order without their prior knowledge and the responses were recorded. So, in total 6 raters assessed a total of 90 clips. As mentioned above, the objective is to analyze thin slices of behavioral responses. Therefore, the extracted clips from the SEMAINE corpus are curtailed on average to 14.09s. The data has 45.5% male and 54.5% female population, interacting with the virtual characters. Figure 1 illustrates the distribution of scores of personality dimensions observed by the raters across different recordings for 9 different dimensions. Overall, almost all of the dimensions are rated high on or near the average score of the personality trait assessed.

IV. Correlations in Trait Scores

As mentioned earlier, the first contribution of our work is investigating how various trait dimensions are inter-related. These relationships, once understood, can be leveraged to learn and train better models towards building an automatic framework. To this aim, for every clip rated, for audio-visual scores and visual-only scores separately, an average score for each dimension is computed. The Figures 2(a) and 2(b) represent the correlation of nine dimensions from audio-visual and visual-only scores respectively. Table I gives a better picture of the findings displaying highest positive and highest negative correlations found. All correlation coefficients are computed using Spearman rank-order correlation, and all correlation values are significant for p-value ($p<<0.001$). Interestingly, communication channel does not appear to affect the correlation between the dimensions across the data set. More specifically, same set of dimensions are found to be highly correlated, either positively or negatively, regardless of the rated clip being audio-visual or visual-only. Looking at

(a) High Positive Correlations.

Dimension 1	Dimension 2	Audio-visual Scores	Visual-only Scores
Facial Attractiveness	Likability	0.8910	0.8749
Facial Attractiveness	Agreeableness	0.7841	0.7056
Facial Attractiveness	Openness	0.7457	0.6414
Agreeableness	Openness	0.6873	0.7616
Agreeableness	Likability	0.8142	0.8765
Openness	Likability	0.8273	0.8039
Extraversion	Engagement	0.7139	0.8084

(b) High Negative Correlations.

Dimension 1	Dimension 2	Audio-visual Scores	Visual-only Scores
Neuroticism	Facial Attractiveness	-0.7577	-0.6694
Neuroticism	Agreeableness	-0.6301	-0.8241
Neuroticism	Likability	-0.6925	-0.7926

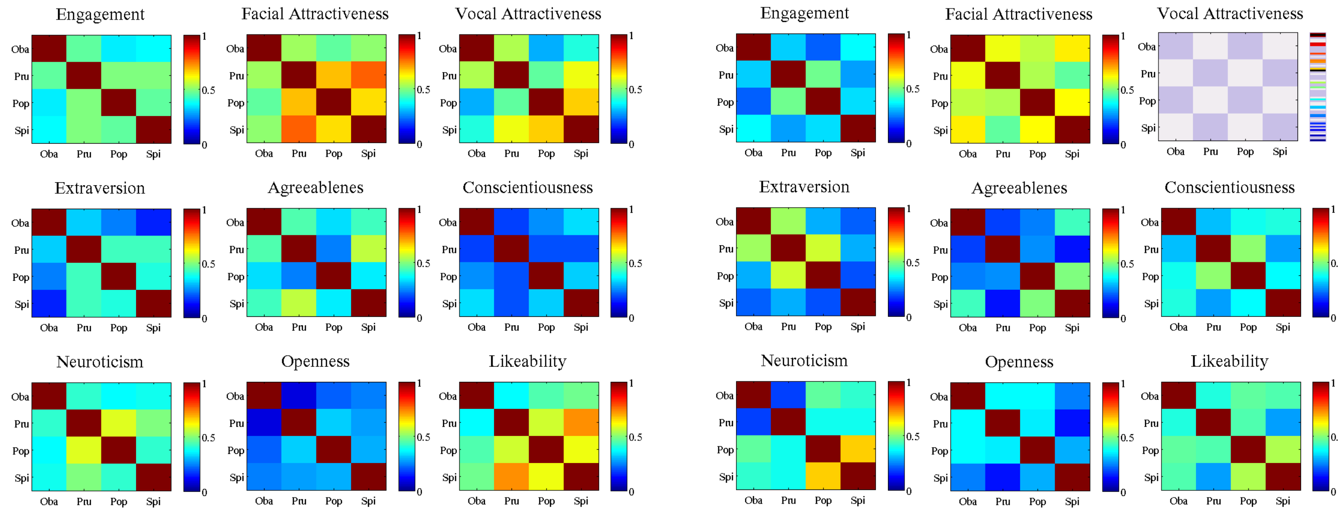
TABLE I. List of dimensions with high correlation.

Table I, *Facial Attractiveness - Likability* and *Facial Attractiveness - Agreeableness* are the two most highly correlated sets of dimensions in both audio-visual and visual-only mode. Negative correlation is found to be strongest between *Facial Attractiveness* and *Neuroticism* for the audio-visual mode. The strength of this correlation decreases slightly in visual-only mode.

Further analysis is done for each dimension in order to explore consensus over the scores given by different raters. Scores for each (audio-visual and visual-only) clip are segregated for every dimension from different raters. Every rater's score for each dimension is then correlated with every other rater's score for different channels. Despite our previous findings that some of the dimensions are highly correlated with each other (see Table I), in this case no strong correlation was found among raters' score for any specific trait dimension. In visual-only scores, there is some positive correlation for the *Facial Attractiveness* dimension between raters 3 & 6 and raters 5 & 6, however, not strong. In audio-visual results, the scores are very much subjective and have no strong correlation among any of the raters for any dimension. This points to the issue reported by Mohammadi *et al.* in [7], where they found low agreement amongst raters. Low correlation between the scores given by different raters suggests that perception of personality trait along a pre-defined scale may indeed be subjective. Every individual may perceive the social attributes differently. As automatic analyzers and predictors depend on labels provided by human raters, this poses a great challenge for creating a model addressing the issue of rater subjectivity and bias.

V. Context and Personality Perception

This section explores whether there are deviations in trait perception and ratings when there is a change of situational context. In other words, we are interested in exploring the change in the perception marked by an external observer when the same individual interacts with different virtual characters manifesting varied social attributes. As discussed in Section III, there are four different virtual characters each with a distinct and strong emotional attribute. To perform this experiment, we isolated the scores rated by 6 different raters for each dimension for every context in the data set. Then we correlated the scores of each dimension observed in 4 different contexts. The aim was to understand whether there are any relationships between the ratings provided for a dimension irrespective of



(a) Audio-Visual Mode (b) Visual-only Mode

Fig. 3. Plots illustrating the correlations for dimension ratings in different contexts. The ticks on the axis in all plots represent different contexts. The color of the blocks represent the strength of correlation as per the legend markings. From left-right and top-bottom: Obadiah (Oba), sad; Prudence (Pru), even-tempered; Poppy (pop), happy; Spike (Spi), angry.

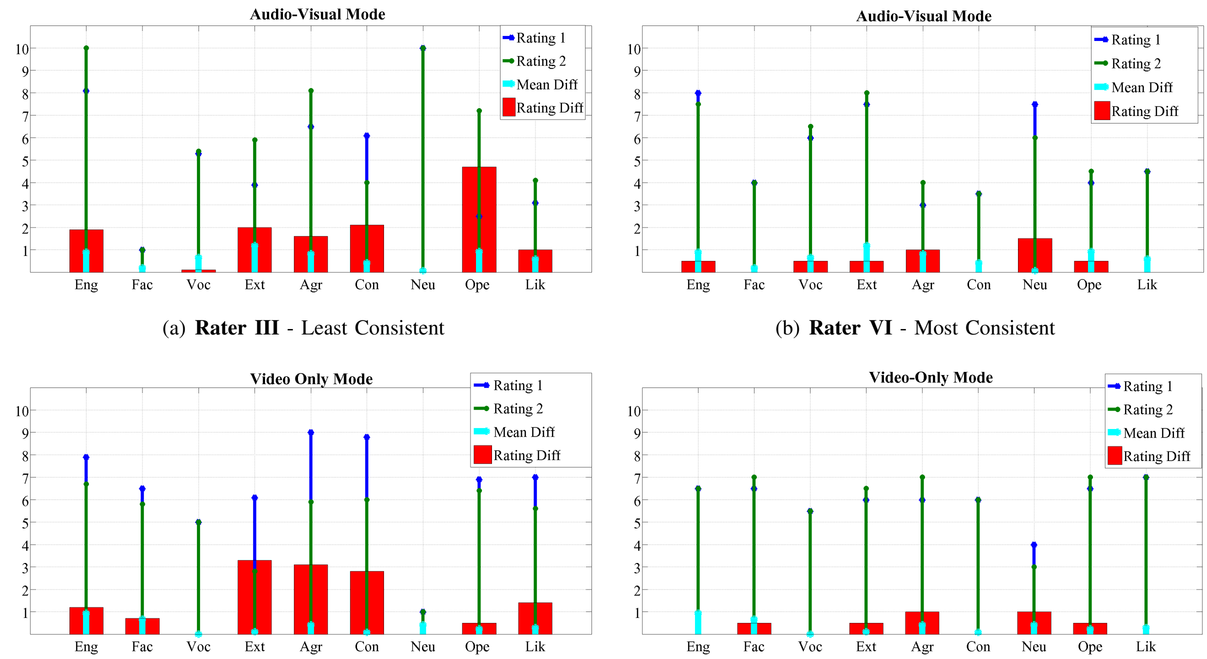
the situational context of the interaction taking place, and to what extent the ratings differ with respect to the situational context.

The plots in the Figures 3(a) and 3(b) represent the correlation of contexts for various dimensions for audio-visual and visual-only modes, respectively. It is evident from the figures that there are only a small number of dimensions showing high positive correlation in both modes. In other words, for most of the trait dimensions context appears to play an important role in altering the raters' perceptions and their scores. These results are in accordance with our hypothesis that with change in the situational and interactional context, participants exhibit different aspects of their trait attributes. Yet, there seem to be a few exceptions, looking at Fig. 3(a), *e.g. Facial attractiveness* shows high positive correlation with correlation coefficient 0.67 and 0.77 for (Pru, Pop) and (Pru, Spi) combination. Now, context Pru is interaction with Prudence (Pru) character who is emotionally sensible and even-tempered and context Pop is interaction with Poppy (Pop) who is a happy character. The interactions with both of these characters generally are well-received, and generate similar facial response, appreciated by the raters. Thus, we observe a strong correlation over the *Facial attractiveness* dimension for these two similar contexts. However, the second case is even more interesting where one context is interaction with Prudence (Pru), and the other context is interaction with Spike (Spi), the angry and the confrontational one. This can be attributed to the fact that with the availability of the audio modality, the external observers are able to listen to the conversation and its content when assessing *Facial attractiveness*. The angry and unfriendly attribute manifested by Spike helps justifying the raters' scores provided for the participants' facial response. Hence, the ratings for *Facial attractiveness* do not appear to drop even when the participant in the video is interacting with the angry character. However, in case of the visual-only mode, the raters have no idea about the conversation taking place, and the *Facial attractiveness* is judged only by the facial responses of the participant. Consequently the correlation between *Facial attractiveness* and the Spike situational context decreases for the visual-only mode (Fig. 3(b)). This also appears to be the case for the Prudence situational context. Similarly to *Facial attractiveness*, *Likability* dimension is also found to exhibit some correlation. However, for all other dimensions the correlation values are overall low. This finding suggests that changes in situational context cause changes in trait perception. Interestingly, there is no negative correlation found for any of the dimension in similar or dissimilar contexts. In Fig. 3(b), dimension *Vocal attractiveness* can be ignored as it corresponds to results obtained in visual-only mode.

VI. Rater Consistency

Another experiment was designed to explore the consistency of the raters in observing the trait attributes discussed in Section IV. The objective was to assess the variation in an individual's rating for different personality traits. As mentioned earlier, all 6 raters were shown a pre-selected audio-visual clip and a visual-only clip twice, but in random order without their prior knowledge. The blue and green markings in Fig. 4 depict the ratings provided by the rater for the first and the second instance respectively. The red bar shows the difference in the two observed ratings. The cyan mark represents the mean difference of a particular trait evaluated by taking the average of differences found in observations made by all the raters for a specific dimension.

The consistency of a rater is based on the difference between the deviation of their ratings for a dimension and the mean difference for that dimension. As evident from Fig. 4, rater VI is the most consistent rater and has shown least deviation in her/his observations for the same clip in both audio-visual and visual-only modes. In audio-visual mode, the individual deviations of rater VI for almost all dimensions are less than the mean difference except for the dimension *Agreeableness* where the deviation is marginally higher than the mean and *Neuroticism* where there is some noticeable difference. In visual-only mode, for the same rater VI, *Agreeableness* and *Neuroticism*, *Extraversion* and *Openness* have shown



(a) **Rater III** - Least Consistent

(b) **Rater VI** - Most Consistent

(c) **Rater III** - Least Consistent

(d) **Rater VI** - Most Consistent

Fig. 4. Representation of rater consistency. The blue and green markings describe the observation provided by the rater for the first and the second instance respectively. The red bar shows the difference in the two observed ratings. The cyan mark represents the mean difference of a particular trait evaluated by taking the average of differences found in observations made by all the raters for a specific dimension.

slightly higher difference between the individual deviation and the mean. However, compared to other raters, rater VI performs best. On the other hand, rater III is the most inconsistent rater with maximum deviations recorded from the mean differences for all dimensions except for *Neuroticism* in both audio-visual and visual-only modes. The result for the third dimension, *Vocal attractiveness*, in the graph 4(c) and 4(d) for the video only mode can be ignored as there was no audio channel to assess this trait.

In audio-visual results from Figures 4(a) and 4(b), *Facial attractiveness*, *Neuroticism* and *Conscientiousness* have the least mean differences among other dimensions and *Openness* and *Extraversion* have the highest mean difference. In visual-only scores, on the contrary to audio-visual findings, *Extraversion*, *Conscientiousness* and *Openness* show the least mean difference. Looking at the performance of all raters in terms of consistency, the ratings are more consistent for visual-only mode. The mean difference for various dimensions is also found to be lower in the case of visual-only mode. This can be attributed to (i) cognitive overload that may affect the performance of a rater while audio-visual feed is processed, and (ii) individual differences in modality preference during perception.

The issue of subjective labeling and low agreement amongst raters, as discussed thus far, is still unresolved. To address this, the credibility of a rater is evaluated by assigning weights to every rater based on their consistency in assessing the same clip. The weight ψ_i for a rater i is calculated as:

$$\psi_{\mathbf{i}} = 1 - \alpha||\delta_i|| \tag{1}$$

$$\alpha = 1/\delta_{max} \tag{2}$$

where δ_i in Equation 1 is the deviation observed in a rater's scores for the same clip and α is computed to adjust for maximum erroneous score of all the raters as shown in Equation 2. The weights are further used to create a weighted model discussed in section VII. It can be argued that such a system may induce bias by penalizing some raters' scores and preferring others. However, the whole point is to identify and segregate such observations which are more likely to induce errors while learning and training a model. Hence, the system is designed to equilibrate these error-prone ratings.

VII. Automatic Personality Prediction

A framework encapsulating a weighted model to account for credibility of a rater for automatic prediction of perceived traits using low-level visual features is described in this section. The face registration was performed using an off-the-shelf face registration and tracking tool, IntraFace [18]. The localized fiducial points are further used for face alignment by computing an affine transform. Pyramid of Histogram of Gradient (PHOG) [19] features are computed on the aligned faces. PHOG is an extension of the popular Histogram of Gradient descriptor. For computing PHOG, an image is divided into blocks on various pyramid levels and histograms are computed based on orientations. Orientations are fused at block and pyramid level into histograms. This approach represents an image by both its local shape, the individual histograms calculated per block, and its spatial layout, the result of multiple resolution tiling [19]. The motivation behind using PHOG is based on its superior performance for face analysis as compared to Local Binary Patterns (LBP) [20].

For training the models for each dimension, linear Support Vector Regression (SVR) [21] is used. A leave-one-subject-out

Dimension	Average Model	Weighted Model
Engagement	0.892	0.544
Facial Attractiveness	1.269	0.716
Extraversion	1.157	0.758
Agreeableness	0.934	0.599
Conscientiousness	0.659	0.405
Neuroticism	1.389	0.974
Openness	0.815	0.569
Likability	1.089	0.737
Average Error Rate	1.025	0.662

TABLE II. THIS TABLE COMPARES THE RMSE OF THE AVERAGE MODEL AND THE WEIGHTED MODEL FOR EACH DIMENSION.

validation approach is followed to evaluate the performance of the framework. As there are 6 different raters evaluating the videos clips, for the average model the ground truth is computed as a mean over the 6 different ratings for each dimension. For the weighted model, separate models are created for different raters using their labels as the ground truth and and a weighted ground truth is obtained using the weight ψ_i defining the credibility of the raters from equation 1.

The original video frame resolution is 720×420 pixels. The frame rate of the video clips is 29 fps. For a video, fiducial point detection and tracking is performed using IntraFace software [18]. The obtained 49 fiducial points are further used to align the faces by computing affine transform and the face blob size is set to 128×128 pixels. PHOG is computed on the aligned face for every frame. The mean and standard deviation of PHOG for the entire video sequence is used as the final feature. First, an average model for each dimension is learned using mean of all the ratings provided by different raters. The results are listed in Table II, where the second column represents the Root Mean Square Error (RMSE) for predicting each dimension using the average model. Further, a weighted model is trained using ψ_i as weights for each rater. The results listed in last column of Table II clearly demonstrate the improvement achieved by the weighted model. On average, error rate of the weighted model is less (0.66) compared to the error rate of the average model (1.02). The weighted model outperforms the average model by automatically predicting each trait dimension more accurately.

VIII. SUMMARY AND CONCLUSIONS

There is an increasing interest in HCI domain to develop smart systems which can understand the various affective and emotional attributes of humans to make interactions with machines more social and believable. In this work, we analysed various aspects of personality prediction in human-machine interaction settings assessed by external observers and raters. First, the relationship between various personality traits was studied. Results showed that some of the dimensions are highly correlated, positively and negatively. As a future work, we would like to take advantage of this finding and incorporate it into our automatic prediction framework. Secondly, we proposed to measure the credibility of the external raters in order to find a solution for errors induced by subjective biases. Results demonstrated that a better model can be learned by separating the consistent and more reliable raters from the inconsistent ones. The impact of context on the perception of the traits was also investigated. The experimental results illustrated that interacting with virtual characters manifesting strong emotional attributes affects the trait perception of the external observers. Lastly, integrating the findings from the above mentioned experiments, we proposed a weighted automatic trait prediction framework using the visual channel. This model outperformed the average model by automatically predicting the trait dimensions more accurately. As future work we will incorporate audio features in the proposed weighted framework.

IX. ACKNOWLEDGEMENTS

The work of H. Gunes is supported by the EPSRC MAP-TRAITS Project (Grant Ref: EP/K017500/1).

REFERENCES

[1] J. S. Uleman, S. Adil Saribay, and C. M. Gonzalez, "Spontaneous inferences, implicit impressions, and implicit theories," *Annu. Rev. Psychol.*, 2008.

[2] T. W. Bickmore and R. W. Picard, "Establishing and maintaining long-term human-computer relationships," *ACM TOCHI*, 2005.

[3] A. Ortony, "On making believable emotional agents believable," *Trappl et al.(Eds.)*, 2002.

[4] E. André, M. Klesen, P. Gebhard, S. Allen, and T. Rist, "Integrating models of personality and emotions into lifelike characters," in *Affective interactions*, 2000.

[5] B. Lepri, J. Staiano, G. Rigato, K. Kalimeri, A. Finnerty, F. Pianesi, N. Sebe, and A. Pentland, "The sociometric badges corpus: A multilevel behavioral dataset for social behavior in complex organizations," in *Privacy, Security, Risk and Trust (PASSAT), SocialComm*, 2012.

[6] F. Mairesse and M. Walker, "Automatic recognition of personality in conversation," in *Human Language Technology Conference of the NAACL*, 2006.

[7] G. Mohammadi and A. Vinciarelli, "Automatic personality perception: Prediction of trait attribution based on prosodic features," *TAC*, 2012.

[8] F. Pianesi, N. Mana, A. Cappelletti, B. Lepri, and M. Zancanaro, "Multimodal recognition of personality traits in social interactions," in *ACM ICMI*, 2008.

[9] O. P. John and S. Srivastava, "The big five trait taxonomy: History, measurement, and theoretical perspectives," *Handbook of personality: Theory and research*, 1999.

[10] L. Jackson, J. E. Hunter, and N. H. Hodge, "Physical Attractiveness and Intellectual Competence: A Meta-analytic Review," *Social Psychology Quarterly*, 1995.

[11] J.-I. Biel, L. Teijeiro-Mosquera, and D. Gatica-Perez, "Facetube: predicting personality from facial expressions of emotion in online conversational video," in *ACM ICMI*, 2012.

[12] S. Argamon, S. Dhawle, M. Koppel, and J. W. Pennebaker, "Lexical predictors of personality type," in *Joint Annual meeting of the interface and the classification society of North America*, 2005.

[13] J. Oberlander and S. Nowson, "Whose thumb is it anyway?: classifying author personality from weblog text," in *The COLING/ACL*, 2006.

[14] A. Vinciarelli, H. Salamin, A. Polychroniou, G. Mohammadi, and A. Origlia, "From nonverbal cues to perception: personality and social attractiveness," in *ICCBS*, 2011.

[15] L. Batrinca, B. Lepri, N. Mana, and F. Pianesi, "Multimodal recognition of personality traits in human-computer collaborative tasks," in *ACM ICMI*, 2012.

[16] J. C. Biesanz and S. G. West, "Personality coherence: Moderating self–other profile agreement and profile consensus." *Journal of personality and social psychology*, 2000.

[17] G. McKeown, M. F. Valstar, R. Cowie, and M. Pantic, "The semaine corpus of emotionally coloured character interactions," in *ICME*, 2010.

[18] Xuehan-Xiong and F. De la Torre, "Supervised descent method and its application to face alignment," in *IEEE CVPR*, 2013.

[19] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in *ICIVR*, 2007.

[20] A. Dhall, A. Asthana, R. Goecke, and T. Gedeon, "Emotion recognition using PHOG and LPQ features," in *IEEE AFGR FERA*, 2011.

[21] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," 2001, http://www.csie.ntu.edu.tw/ cjlin/libsvm.