Predicting personality traits from touchscreen based interactions

Ludwig Küster, Carola Trahms, and Jan-Niklas Voigt-Antons
Quality and Usability Lab
Technische Universität Berlin
Email: ludwig.kuester@posteo.de, carola.trahms@tu-berlin.de, jv@qu.tu-berlin.de

***Abstract*—Human influence factors can have a strong impact on how users perceive the quality of a given system. Traditional measures such as questionnaires can be a time consuming and partially invasive means to assess human influence factors like personality. In this paper, we investigate whether an individual's personality traits can be classified based solely upon the characteristics of their touchscreen usage. We record the tablet input of 75 subjects using a cognitive training application. The application requires the user to spell words by tapping on letters and dragging them to their denoted position on the screen. The task is presented in two conditions: in one with normal functionality of the program, the other with the usability of the application impaired. The subject's personality is measured with the NEO-FFI questionnaire and captured in the five dimensions of the five-factor model (Big 5). The personality scores and 68 features (touch-behavior and task-performance), as well as statistical metrics derived from them, are fed to a number of classification algorithms. Our results show that in a binary classification task with normal usability, high levels of 'Neuroticism' can be distinguished from low levels with a mean accuracy of 66 percent. The extent to which a personality dimension can be predicted is dependent on the usability of the test system. In a normal setting, 'Extraversion' can be predicted with an accuracy of 61 percent. However, with impaired usability, the prediction rises to an average of 67 percent. The experiment shows that records of touchscreen interactions allow for the prediction of personalities significantly better than random. The study of the human influence factor personality and its relation to perceived quality would be facilitated by using touchscreen interaction data as a fast and easily accessible estimate of a user's personality.**

I. INTRODUCTION

Personality is a pivotal element in the understanding of human behavior. In the human computer interaction field, the pursuit to comprehend the influence of personality on the user perception has manifested in a surge of publications [1]. Firstly, modeling the user's characteristics opens new possibilities for tailoring content for individual needs. Secondly, quality of experience is known to be user dependent [2], [3]. Examples include the dependency of the perceived technical audio quality on listener types [4], and the correlation of personality traits with the usability ratings of products [5]. Personality has stood out as a predictor. It can forecast peoples learning approaches, their academic success, their intelligence or ratings of subjective well being [6]. The connection of personality psychology to the human-computer interaction field was fueled by the vision of training machines with social and affective intelligence, and facilitated by the numerical nature of personality psychology's most influential paradigm [7], the five factor model (Big-5) [1]. Personality psychology builds on the assumption that individual characteristics manifest in stable and invariant behavioral patterns which can be captured in quantitative terms [8]. In the past, different data sources have been used to infer the personality dimensions. The traditional methods are self-assessment questionnaires [9], which have been in recent years complemented by work estimating personality from social media [10] or even phone-bills [11]. While reaching high accuracies, these methods require comprehensive access to personal information and long data sampling periods. By contrast, touchscreen data can easily be obtained by smartphone or tablet applications. First approaches predict user experience from touch interactions [12]. To date, there is no publication of personality prediction from the touch modality.

The second section of this paper introduces research from adjacent fields. The third section presents our experiment and its conditions. We first classify the personalities of 75 participants for a test system with normal usability, then compare it to a setting of impaired usability. Finally we present and discuss our results.

II. RELATED WORK

The five-factor model, commonly referred to as the Big-5, is a metric of personality traits and the dominant paradigm in personality research today [7]. It builds on the notion that adjectives, which people use to describe each other and themselves, have semantic similarities and relationships [13]. The human judgment on these semantic similarities groups the adjectives and thus reduces them to a few underlying dimensions [1]. Continuous bipolar scales represent these dimensions. In the five-factor model they encompass: openness, conscientiousness, extraversion, agreeableness and neuroticism (OCEAN). For example, people with high levels of openness would be described as imaginative or interested, whereas low levels are associated with practical or conforming behavior. The most common instrument for the assessment of the five dimensions are self-assesment questionnaires with Likert scales. Here, participants rate their own behavior [9]. Personality could alternatively be captured in informant reports, behavioral measures or a combination of methods.

The emotional state of a user is influenced by personality and can be affected by the usability of the used system [14]. The usability can be controlled by manipulating selected

978-1-5386-2605-4/18/$31.00 ©2018 IEEE

properties of a system. Well designed applications lead to more positive subjective feelings. Different personality types react with different intensity to a less usable application [5]. The inclination to show positive emotions can be strongly predicted by extraversion, while neuroticism is a predictor of negative emotions [15]. Hence the ability to predict different personality dimensions based on touch data might be influenced by the usability of the test system. The notion that a touchscreen task in which participants react to visual stimuli might convey patterns that link to personality is further supported by evidence that people with low levels of extraversion (introverts) are faster in processing sensory input into motor output [16]. We explored three adjacent fields to gather best-practice approaches for designing the experiment and feature collection. A first line of research explores the relation between phone or smartphone data to the five-factor model. Smartphone data or phone logs offer an insight into human communication and movement patterns, with unprecedented scope and depth. Secondly, there is a rich body of work attempting to disclose users' emotional states from behavioral information. Among various data sources, touchscreen interactions have been successfully analyzed and used to inference the user's emotions. This field offers insights into both the relation of touchscreen-interaction to emotion, and the interplay between emotion and personality. In the third area, user-authentication researchers have investigated the possibilities of turning touchscreen behavior into statistical features. Their security related aim is securing systems by mapping characteristic touch patterns onto individuals. This builds on exploiting the entropy of touch actions and serves as a source of inspiration for the construction of metrics.

A multitude of work has used different kinds smartphone data for the prediction of user's personalities. These include data from a) social interaction (phone usage), b) proximity (via Bluetooth), c) information about which apps were used on the phone and d) phone logs (itemized bills) [11], [17], [18]. In a 2011 study Chittaranjan et al. [18] realized that features obtained from smartphone usage can be indicators of the Big-5 traits. In classifications, the median of each of the five personality dimension was used to split the target data into two classes. Their classification reached its highest accuracy for openness (59%) [18]. This work was extended in 2013 by Montojoye et al. who computed a novel set of indicators based on mobile-phone logs made available by phone-companies [11]. These logs reveal the number and frequency of phone interactions with others, as well as phone owners' locations and movement patterns, derived from the logging of the network. They divided every dimension of the Big-5 into three classes. In a classification task, their prediction reached a mean accuracy across traits of 61%. In 2016 Monsted et al. collected data from the smartphones of 730 individuals [17]. They separated each personality dimension into three target classes and reached prediction accuracies 11% above the random chance baseline. They declare that only extraversion and to some degree, neuroticism, can be predicted by smartphone usage patterns. Extraversion is known to be a strong predictor of positive emotions, whereas neuroticism predicts negative ones [15]. This relation of personality to emotions makes it promising to apply techniques from emotion analysis to personality prediction. Studies capture swipes and taps. They calculate the frequency of touches, or the deviation from a theoretical optimal line connecting start and end point of a swipe [19]. Gao et al. define four emotional states (excited, relaxed, frustrated and bored) and differentiate between them with 69% to 77% accuracy. The pressure applied to the screen is highlighted as an especially informative feature [20].

Some user authentication research has turned its focus onto touchscreen behavior. It assumes that touch-characteristics can be regarded as an invariant feature of human behavior. With touch patterns as a biometric feature, touchscreen metrics should be used to ascertain that a device is only used by its owner. These works have placed emphasis on extracting the maximum entropy from touch actions with a sophisticated set of statistical features. Variables are extracted from smartphones' touchscreen logs and data, storing the complete swipe trajectory. This allows for the computation of 30 behavioral metrics, including the speed at the start and end of each swipe, or its directness [21]. A user's tapping patterns might display unique characteristics, as they are different in strength, rhythm and angle of the finger movements. Touchscreen inputs are supplemented by data from other smartphones sensors, such as the accelerometer [22]. In sum, the related work reveals that the prediction quality heavily relies on a set of meaningful input variables. A smart aggregation of touch-screen raw-data into higher-level metrics promises high classification scores.

III. METHODS

We recorded the touchscreen interactions of test participants in a controlled experiment. The data was fed into classification algorithms to predict users' personalties based upon their touch behavior.

Participants were invited to the lab and asked to play the cognitive training game Spell on a tablet device. The game asks the user to spell a word displayed on the bottom side of the screen, with its letters randomly scattered around the screen. The player must tap on the sought letter and drag it down to its position denoted by a place holder. The game continues with a new word, after the initial word is spelled correctly. This experiment design captures both taps and swipes in a controlled setting. Personality dimensions are strong predictors of how an individual shows positive or negative emotions. Also, a person's emotions while interacting with a system depend on the system's usability. We therefore conducted an experiment with two conditions: Participants would first perform the Spell game in a normal setting. Second, they would play a version of the same application, this time with extremely small interface elements (icons). In this mode of impaired usability, the user is challenged with not being able to tap and drag the letters as easily. The task of spelling therefore becomes a challenge. This application design allowed us to subsequently examine if the prediction quality for personality is different once a system's usability is impaired. During the experiment, the participants were asked to use the application with their dominant hand and play several 2-4 minute sessions. The experiment is done in randomized order design, with each condition appearing twice. Subjects then filled in a paper version of the NEO-FFI questionnaire, which comprises of 60 statements. The degree of agreement or disagreement concerning a statement is rated by the subject on a 5-point Likert scale. Each of the 5 OCEAN dimensions has 12 statements associated [23]. An individual's age and gender significantly influences the Big-5 score. Women score higher than men in all 5 scales of the NEO -FFI [24]. In addition, all five scales correlate significantly with age. Older people tend to be less neurotic, less open and less extroverted while having higher values of agreeableness and conscientiousness [23]. The handedness variable is included as differences in hand geometry are known to influence touch-characteristics [25].

A participant's touchscreen interactions during use of the test application are captured in great detail and stored on a server. The Spell game, which is used as the test software, is part of the TU-Berlin developed PflegeTab application which seeks to improve the quality of life for people with dementia [26]. On one hand, the interaction records capture a players performance in the "Spell" task. Examples are the number of correct drag-and-drop events, touches outside of the answer area or the time a subject spent on a task. On the other hand, the interaction record stores the general way a user touches the screen. These touch events are stored with their x-y coordinates and time-stamps of every finger-down and finger-up movement. One minute of application-usage provides us with around 100 performance or touch events being stored on the server. The success of a later classification task heavily depends on the transformation of this mass of raw-data into a meaningful set of features. We aggregate performance events into performance variables such as counts of successfully placed letters, false drag-and-drop actions or spelling errors.

One side of the feature set describes the performance during the "Spell" game. The other side describes the general way a user touches the screen, the touch variables. One variable on this side of the feature set is the time difference between a finger down and a finger up event, called the touch-duration. It can convey information about the touch-pressure, as a fiercer touch also takes more time [21]. Other variables include the length and speed of swipes, the frequency of taps and the touch-accuracy. For all touch variables we compute ten statistical parameters (e.g. sd, max and min).

In sum, server records of performance and touch-behavior are transformed into 110 variables. This feature set includes the variables themselves, as well as their statistical derivatives, which leads to high inter-correlations. A smaller number of variables increases the efficiency of the classification-algorithms and makes the data-set easier to handle. We therefore left out variables with an inter-correlation higher than 80%, leaving 44 variables in the feature set. As this is the first attempt to predict personality from touch, we approach the classification as a binary problem - separating each personality dimension into two target classes. The scoring system of the NEO-FFI questionnaire delivers metric values in the range of 0 to 48 for each dimensions. Answering every questions in a neutral manner would result in a score of 24. However, the empirical means of the norm-populations deviate distinctly from the theoretical mean of 24. For splitting each dimension into two target classes, we use the median of our population as a separation margin. The split is done so that both target classes contain an equal number of observations.

Our data analysis and classification are carried out in the statistics language R. Its "caret" package provides a uniform interface for the data-preparation, the classification, and the conclusive comparison of different classification algorithms [27]. We then train and test different classifiers. A Support Vector Machine (SVM) tries to separate the data points of two target classes by an optimal margin. For higher dimensions however, a hyperplane is unlikely to separate points well. The distance between points and the separation margin is then translated into non-Euclidian space and a radial-basis-function (RBF) used as kernel. Two hyperparameters must be found for this model: the misclassification cost C (the number of points that can deviate from the separation) and the sharpness sigma of the Gaussian basis functions. The Random-Forest (RF) algorithm works well for exploratory modeling and has an implicit feature-selection mechanism that arises from the structure of the classifier. It is tuned by its predictor number. The K-Nearest Neighbor algorithm (KNN) does not require the input variables to follow certain distribution. Its places an unseen observation into the same category, the majority of its nearest neighbors belong. A logistic regression (LR) is suited for binary target data, which allows for a linear separation. It does not make assumptions about the data distribution and is also robust and flexible. As another non-linear model we added Naive Bayes (NB). It is a statistical classifier estimating an observation's membership in target class based on conditional probability. A number of tuning parameters specific to each algorithm can be either automatically optimized or manually defined as a search grid. We used 10-fold cross validation with 10 repeats. This amounts to 100 runs for each single version of an algorithm. The same seed for the random-number generator ensures comparable results between different classifiers and tuning parameters. In all cases the mean-accuracy was used as a target parameter for optimization. Accuracy is the number of true predictions in relation to all predictions made, averaged over all runs of the cross fold-validation. The data from all runs yields a probability distribution for the accuracy. For every mean accuracy value it is tested via the Wilcoxon rank-sum test (alpha-level 5%), whether the observed average accuracy is significantly different from the random chance baseline of a binary classification problem of 50%. The classification is separately done for each of the five personality dimensions. First on the data-set of the normal-usability condition.

Secondly for the experimental session with impaired usability. Two conditions times five separate dimensions amount to 10 classification cases in total. Each of these 10 cases is approached with the set of 5 classifiers (SVM, RF, LR, KNN,

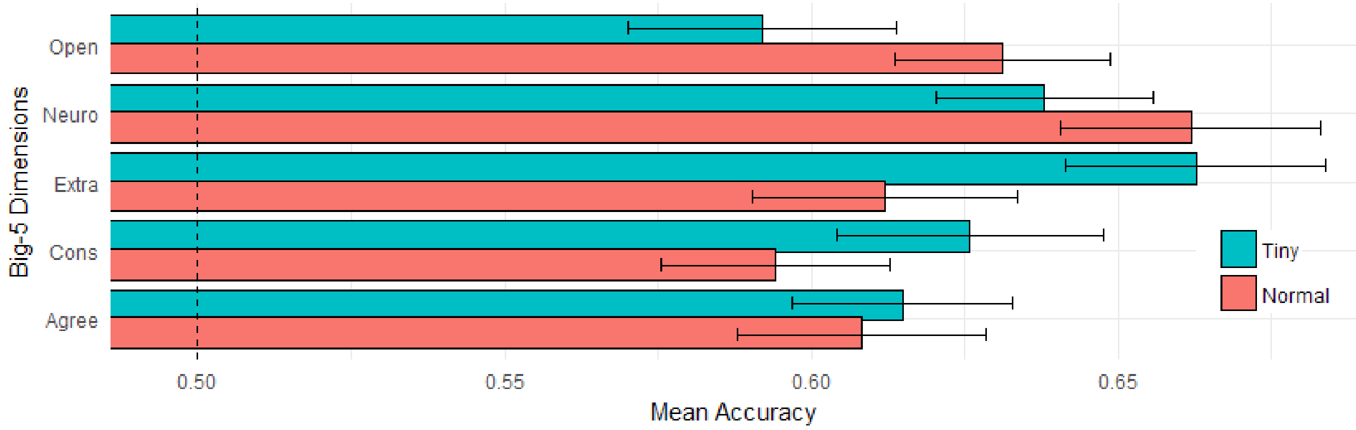

Fig. 1: Mean accuracies of classification with confidence intervals. 10-fold cross validation repeated 10 times. The respective upper bar for each personality dimensions displays the impaired usability (tiny icons) condition. The lower one shows the mean accuracy for a normal usability of the test system. The dashed line is the random chance probability of 50%.

TABLE I: The results of the best performing classifiers for all five personality dimensions. The middle column lists the mean-accuracies with normal usability, the right one for the test condition with impaired usability. Extraversion can be predicted significantly higher with impaired usability. The abbreviation behind the accuracies denote the best performing classifier: Support-Vector-Machine (SVM), Random-Forest (RF) or Logistic Regression (LR).

Dimension	Normal-Condition	Impaired-Usability
Openness	0.65 - SVM	0.59 - SVM
Conscientiousness	0.59 - RF	0.63 - SVM
Extraversion	0.61 - LR	0.67 - RF
Agreeableness	0.61 - SVM	0.62 - SVM
Neuroticism	0.66 - SVM	0.64 - SVM

NB). The performance of these five is compared and the best-performing algorithm selected.

IV. RESULTS

The final data set comprises of records of 75 participants. This includes the features of their touchscreen interactions, their Big-5 personality scores and demographic variables. 36 participants were women, 39 men. They were between 18 and 40 years old, with an average of 27 years. An empirical benchmark for comparing a single sample to the overall population is offered by the NEO-FFI manual, which covers [around then thousand cases. Country and references removed for double blind review]. This norm population is segmented by age group and gender. The bulk of our study participants can be compared to the 21 to 29 year cohort of the norm population. Overall, the means and distributions of our sample are in line with the population means of the respective gender and age cohort. The match of our population-means with the norm-population means supports the use of our population-means as separation margins for the target classes [18].

A total of ten classification cases were examined. Each was approached with a Support-Vector-Machine (SVM), a Random-Forest (RF), a K-Nearest Neighbor (KNN), a Logistic Regression (LR) and a Naive Bayes (NB) classifier. The mean-accuracy for the best performing classifier can be seen in Table 1. A graphical comparison between the cases is provided by Figure 1. In normal condition, openness is predicted by a SVM with 65%, conscientiousness with RF at 59% and extraversion in linear regression with 61%. Agreeableness (61%) and neuroticism (66%) are both best classified with a SVM. Under impaired usability, the prediction of openness reaches 59%, conscientiousness 63%, extraversion 67%, agreeableness 62% and neuroticism 64%. For conscientiousness the best classifiers is now a SVM, for extraversion a Random Forest. Statistical tests for all values prove on an alpha level of 5%, that the difference between the predictions and the random-change baseline of a binary-classification-task of 50% is significant.

The highest prediction value given a normal usability of the test system is reached for neuroticism with 0.66% accuracy, or 16 percentage points better than the baseline. While the prediction of extraversion ranges at 61% given the normal condition, with impaired usability this values rises to 67%, 17 percentage points above the baseline. The whiskers at the end of each bar in Figure 1 denote the 95% confidence interval. They show that an altered usability significantly changes the prediction quality for some personality dimensions. Openness can be predicted 6 percentage points better in normal-condition than with impaired usability. Extraversion however is predicted 6 percentage points better with the test-application displaying tiny-icons. Each classification model yields a ranking of those variables which had the greatest influence on its results. For each of our ten classification cases, we picked the best performing model and collected its respective three variables with the highest prediction value. We could subsequently compare our variables in respect to their entropy. Although the feature set contains both task-related performance variables

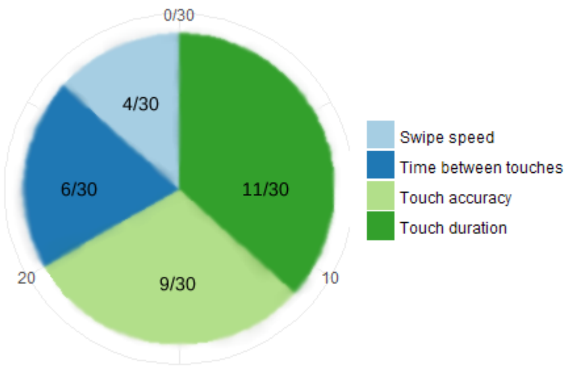

Fig. 2: Four base variables and their statistical derivatives make up the list of the classifications' top predictors. All four variables capture task-independent touch behavior. Touch duration appeared 11 times in the list of the 30 best predictors.

and general touchscreen interaction variables, only the latter appeared among the top three predictors. This holds for all ten classification problems. The list of these three times ten top-predictors comprises of statistical variables (e.g. sd, max). These statistical variables are derivatives of only four base-variables. These most informative base variables are: The mean speed of a swipe, the mean time between two touch events, the mean touch accuracy, and the mean duration of touches. Figure 2 displays the distribution of these four base variables among the 30 top-predictors. Touch duration and accuracy together make up two thirds of this list of the most informative variables.

V. DISCUSSION

To our knowledge our work is the first attempt to derive the Big-5 from touchscreen interactions. A two class prediction problem shows that in all dimensions, personality can be predicted significantly better than random chance, at best 17 percentage points above the baseline. Given a normal usability of the test system, the two dimensions that could be best predicted were neuroticism (66%) and openness (65%). Previous works have found that neuroticism and extraversion are the two dimensions most strongly related to the display of emotions [15]. Monsted et al. [17] observed that especially extraversion and to some extend neuroticism could be predicted from smartphone usage patterns. The version of the test application with impaired usability led to a rise of the prediction accuracy for extraversion by 6 percentage points. While for neuroticism the accuracy decreased by 2 percentage points, neuroticism is still the second most predictable dimension given an impaired usability. It is left to future research why the precition for neuroticism is less dependent on usability. Our predictions reach their highest accuracy for neuroticism (66%) and extraversion (67%). This is in line with previous work. By manipulating selected properties of our test system we infused two conditions of different usability. We have found that the classification differs even under the influence of small impairments such as shrinking the interface elements (icons). The prediction accuracy significantly dependents on the usability of the test application. It is intuitive to think that a hot-tempered person might react more strongly to system flaws. Still, the exact influence of usability on the prediction of personality needs further investigation.

Although there is no touchscreen based prediction to compare our values to, a benchmark for the classification accuracy can be derived from related fields. Taking the usage-data of smartphones as a base for a two class prediction of the five personality dimensions, Chittaranjan et al. [18] reached their highest accuracy for openness (59%). In 2013 Montjoye et al. [11] could predict three classes based on mobile-phone logs, with an across-trait average accuracy of 61%. Other data-sources such as Speech [28] or Social-Networks [10] have proved to be powerful for the prediction of personality. The papers discussed which build on smartphone records require long periods for their data collection. Touchscreen interaction data is collected within a few minutes. If applications could classify their users personality within minutes of use, this would open up new opportunities for content creation and the understanding of personality as a human factor influencing perceived quality. While this paper is the first exploration into predicting personality from touch, we have identified several factors that would lead to a more accurate classification. It has shown that task-independent performance variables lead the ranking of the most informative variables, in all five dimensions. These were swipe speed, time between touches, touch accuracy and touch duration. Touch inputs may be recorded in greater detail by a) storing the complete trajectories of swipes. This captures information such as the deviation from an optimal path between start and end point of a swipe, or the speed at the start and end of each swipe [21]. The touch duration can be regarded as a proxy for the tapping strength, and it has turned out to be one of the most informative variables. Touch duration could be captured by b) the use of other smartphone or tablet sensors such as the accelerometer or receiving the touch-point-size directly from the operating system. The classifier performance could be also boosted by more samples for training. This could either be achieved by recording more participants, or aggregating the touchscreen raw-data over shorter time containers. In this study we averaged events as mean values over one session of the experiment.

Other studies have shown that the five-factor model is not always exhausting the information datasets contain on personality [28]. This can be overcome by clustering personalities by types rather than fixed dimensions and promises better results. Our study was conducted in a laboratory environment, which brings natural limitations to a direct transfer onto real world environments. Users might walk around while using their touchscreen device or face distractions during the interaction. In addition, classifying a user during a free interaction is certainly different to asking for the completion of a cognitive training game. Although the high prediction value of taskunrelated variables might show some dependence on the task itself, we are hopeful that the scope of the prediction methodology can be extended in the future.

VI. CONCLUSION

This paper is a first exploration of the relation between touch and personality. It proves the existence of meaningful and stable connections. Our classification results are significantly higher than random chance baseline of a binary classification task (50%). Extraversion and neurotisicm are the two best predictable personality dimensions. Extraversion is classified 17 percentage points better than random, in a setting with impaired usability. In a normal setting, neuroticism is predicted 16 percentage points better than random. The classification accuracies are comparable to studies with more comprehensive data sets. The preponderance of the related work uses data that requires comprehensive access to personal information and long data sampling periods. The data for this study was recoded during experimental session of only a few minutes. The features most valuable to the classifications were statistical derivatives of only four base variables. All of them characterize task-independent touchscreen behavior, like the speed of swipes or the duration of touches. Firstly, capturing task-independent touch behavior in greater detail promises an increase of the classification accuracy. Secondly, our most valuable predictors are task-independent. This suggests a transfer of our methodology onto non-laboratory and more general use cases. Some personality dimensions, such as extraversion, have proven to be more reliably predicable in a setting of impaired usability. The relation of usability and the prediction quality for personality is yet to be examined. In sum, touchscreen interactions are a promising data source for the recognition of personality. Applications could classify user personalities within minutes - and without the need for comprehensive and private data. This opens up new opportunities for content creation and the understanding of personality as a human factor influencing perceived quality.

ACKNOWLEDGMENT

We thank Giorgio Colombo for his comprehensive work on the experiment.

REFERENCES

[1] A. Vinciarelli and G. Mohammadi, "A survey of personality computing," vol. 5, 05 2014.
[2] S. Möller and A. Raake, *Quality of Experience: Advanced Concepts, Applications and Methods*, 01 2014.
[3] J.-N. Antons, S. Arndt, K. De Moor, and S. Zander, "Impact of perceived quality and other influencing factors on emotional video experience," in *Proc. 7th Int. Workshop on Quality of Multimedia Experience (QoMEX 2015)*. 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331: IEEE, may 2015, pp. 1–6, electronic, online.
[4] T. Walton and M. Evans, "The role of human influence factors on overall listening experience," *Quality and User Experience*, Jan 2018. [Online]. Available: https://doi.org/10.1007/s41233-017-0015-4
[5] P. Kortum and F. L. Oswald, "The impact of personality on the subjective assessment of usability," *International Journal of Human–Computer Interaction*, vol. 1, no. 4, pp. 1–10, 2017.
[6] N. Hayes and S. Joseph, "Big five correlates of three measures of subjective well-being," vol. 34, pp. 723–727, 03 2003.
[7] R. McCrae, "The five-factor model of personality," pp. 1–906, 01 2009.
[8] I. D. G. Matthews and M. Whiteman, "Personality traits," 2009.
[9] G. Boyle and E. Helmes, "Methods of personality assessment." *The Cambridge handbook of personality psychology*, pp. 110–126, 2009. [Online]. Available: https://doi.org/10.1007/s41233-017-0015-4
[10] Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli, and D. Stillwell, "Personality and patterns of facebook usage," in *Proceedings of the 4th Annual ACM Web Science Conference*, ser. WebSci '12, 2012, pp. 24–32. [Online]. Available: http://doi.acm.org/10.1145/2380718.2380722
[11] Y.-A. de Montjoye, J. Quoidbach, F. Robic, and A. S. Pentland, *Predicting Personality Using Novel Mobile Phone-Based Metrics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 48–55. [Online]. Available: https://doi.org/10.1007/978-3-642-37210-0_6
[12] C. Trahms, S. Möller, and J.-N. Voigt-Antons, "Towards prediction of user experience from touch interactions with mobile applications," in *HCI International 2017 – Posters' Extended Abstracts*, C. Stephanidis, Ed. Cham: Springer International Publishing, 2017, pp. 505–512.
[13] L. R. Goldberg, "The structure of phenotypic personality traits." pp. 26–34, 1993.
[14] M. Thüring and S. Mahlke, "Usability, aesthetics and emotions in humantechnology interaction," *International Journal of Psychology*, vol. 42, no. 4, pp. 253–264, 2007. [Online]. Available: http://dx.doi.org/10.1080/00207590701396674
[15] A. Gomez and R. Gomez, "Personality traits of the behavioural approach and inhibition systems: associations with processing of emotional stimuli," *Personality and Individual Differences*, vol. 32, no. 8, pp. 1299 – 1316, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0191886901001192
[16] J. Stahl and T. Rammsayer, "Differences in the transmission of sensory input into motor output between introverts and extraverts: Behavioral and psychophysiological analyses," *Brain and Cognition*, vol. 56, no. 3, pp. 293 – 303, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S027826260400185X
[17] B. Monsted, A. Mollgaard, and J. Mathiesen, "Phone-based metric as a predictor for basic personality traits," *CoRR*, vol. abs/1604.04696, 2016.
[18] G. Chittaranjan, J. Blom, and D. Gatica-Perez, "Mining large-scale smartphone data for personality studies," *Personal Ubiquitous Comput.*, vol. 17, no. 3, pp. 433–450, Mar. 2013. [Online]. Available: http://dx.doi.org/10.1007/s00779-011-0490-1
[19] S. Shah, J. N. Teja, and S. Bhattacharya, "Towards affective touch interaction: predicting mobile user emotion from finger strokes," *Journal of Interaction Science*, Nov 2015. [Online]. Available: https://doi.org/10.1186/s40166-015-0013-z
[20] Y. Gao, N. Bianchi-Berthouze, and H. Meng, "What does touch tell us about emotions in touchscreen-based gameplay?" *ACM Trans. Comput.-Hum. Interact.*, vol. 19, no. 4, pp. 31:1–31:30, Dec. 2012. [Online]. Available: http://doi.acm.org/10.1145/2395131.2395138
[21] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. X. Song, "Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication," *IEEE Transactions on Information Forensics and Security*, vol. 8, pp. 136–148, 2013.
[22] N. Zheng, K. Bai, H. Huang, and H. Wang, "You are how you touch: User verification on smartphones via tapping behaviors," Tech. Rep., 2012.
[23] P. Borkenau and F. Ostendorf, *NEO-FFI : NEO-Fnf-Faktoren-Inventar nach Costa und McCrae, Manual*. Hogrefe, 2008.
[24] Y. Weisberg, C. De Young, and J. Hirsh, "Gender differences in personality across the ten aspects of the big five," *Frontiers in Psychology*, vol. 2, no. AUG, 2011.
[25] C. Holz and P. Baudisch, "Understanding touch," in *29th Annual CHI conference on human factors in computing systems*, D. Tan, G. Fitzpatrick, C. Gutwin, B. Begole, and W. A. Kellogg, Eds. New York: Association for Computing Machinery, 2011, p. 2501.
[26] J.-N. Antons, J. O'Sullivan, S. Arndt, P. Gellert, J. Nordheim, S. Möller, and A. Kuhlmey, "Pflegetab: Enhancing quality of life using a psychosocial internet-based intervention for residential dementia care," in *Proceedings of International Society for Research on Internet Interventions*. Sanford NC, USA: International Society for Research on Internet Interventions, 2016, pp. 1–1, online.
[27] M. Kuhn, "Building predictive models in r using the caret package," *Journal of Statistical Software, Articles*, vol. 28, no. 5, pp. 1–26, 2008. [Online]. Available: https://www.jstatsoft.org/v028/i05
[28] T. Polzehl, *Personality in Speech: Assessment and Automatic Classification*. Springer Publishing Company, Incorporated, 2014.