A Survey of Personality Computing

Alessandro Vinciarelli *Member, IEEE*, and Gelareh Mohammadi

Abstract—Personality is a psychological construct aimed at explaining the wide variety of human behaviors in terms of a few, stable and measurable individual characteristics. In this respect, any technology involving understanding, prediction and synthesis of human behavior is likely to benefit from Personality Computing approaches, i.e. from technologies capable of dealing with human personality. This paper is a survey of such technologies and it aims at providing not only a solid knowledge base about the state-of-the-art, but also a conceptual model underlying the three main problems addressed in the literature, namely Automatic Personality Recognition (inference of the true personality of an individual from behavioral evidence), Automatic Personality Perception (inference of personality others attribute to an individual based on her observable behavior) and Automatic Personality Synthesis (generation of artificial personalities via embodied agents). Furthermore, the article highlights the issues still open in the field and identifies potential application areas.

Index Terms—Personality, Automatic Personality Perception, Automatic Personality Recognition, Automatic Personality Synthesis

1 INTRODUCTION

It is at least since the times of the Greek philosopher Theophrastus (c. 371 - c. 287 BC) that individual differences are the subject of scientific inquiry: "*I applied my thoughts to the puzzling question - one, probably, which will puzzle me for ever - why it is that, while all Greece lies under the same sky and all the Greeks are educated alike, it has befallen us to have characters so variously constituted*" [1]. Personality psychology is the modern answer to such an ancient question: As a construct, personality aims at capturing stable individual characteristics, typically measurable in quantitative terms, that explain and predict observable behavioral differences [2].

Current personality models succesfully predict "*patterns of thought, emotion, and behavior*" [3] as well as important life aspects, including "*happiness, physical and psychological health, [...] quality of relationships with peers, family, and romantic others [...] occupational choice, satisfaction, and performance, [...] community involvement, criminal activity, and political ideology*" [4]. Furthermore, attitude and social behavior towards a given individual depend, to a significant extent, on the personality impression others develop about her [5].

Such an effectiveness in capturing the crucial aspects of an individual is probably the main reason behind the interest of the computing community for personality. Figure 1 shows the number of papers including the word "*personality*" in the title on *IEEE Xplore* and *ACM Digital Library*, probably the two most important repositories of computing oriented literature. While being only the tip of the iceberg - most articles revolving around personality do not mention it in the title[1] - these papers clearly show that the interest for the topic is growing and that the trend promises to continue in the foreseeable future.

After the earliest, pioneering approaches aimed at integrating personality psychology in Human-Computer Interaction [6], the interest for the topic was fueled by three main phenomena in the technological landscape. The first is the increasing amount of personal information, often self-disclosing beyond intention [7], available on social networking platforms [8]. The second is the possibility of collecting everyday spontaneous, fine-grained behavioral evidence through mobile technologies and, in particular, smartphones [9]. The third is the attempt of endowing machines with social and affective intelligence, the ability of interacting with humans like humans do [10]. The three phenomena are probably the reason of the sudden rise of interest for the topic in the mid 2000s (see Figure 1).

Overall, personality is relevant to any computing area involving understanding, prediction or synthesis of human behavior. Still, while being different and diverse in terms of data, technologies and methodologies, all computing domains concerned with personality consider the same three main problems, namely the recognition of the true personality of an individual (Automatic Personality Recognition), the prediction of the personality others attribute to a given individual (Automatic Personality Perception), and the generation of artificial personalities through embodied agents (Automatic Personality Synthesis). To the best of our knowledge, this is the first survey of the approaches addressing these problems. The main works presented in the literature are analyzed in terms of type of behavioral evidence they adopt, amount of data and subjects they use, actual tasks they propose and performance they achieve. Furthermore, the article surveys psychological work aimed at establishing a link

- *A.Vinciarelli is with University of of Glasgow and Idiap Research Institute. E-mail: vincia@dcs.gla.ac.uk*
- *Gelareh Mohammadi is with Idiap Research Institute and EPFL. E-mail: gelareh.mohammadi@idiap.ch*

1. At the moment this survey is being written, the query "*personality*" (restricted to articles published since 2000) returns 6172 and 1033 hits on *ACM Digital Library* and *IEEE Xplore*, respectively.

between personality and use of computing technologies, one of the main starting points for the development of personality computing approaches.

The rest of this paper is organized as follows: Section 2 introduces the concept of personality (with particular emphasis on trait based models and the Big Five) and the techniques for its measurement. Section 3 introduces the Brunswik Lens and shows how this latter encompasses the problems addressed in Personality Computing. Sections 4, 5 and 6 survey the approaches on Automatic Recognition, Perception and Synthesis of personality, respectively. Section 7 aims at outlining a research agenda by showing some of the most important issues still open and, finally, Section 8 draws some conclusions.

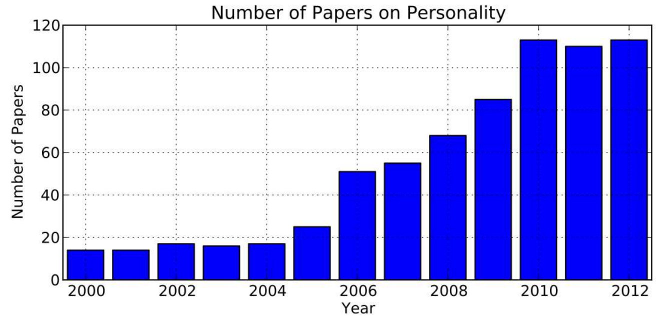

Fig. 1. The chart reports the number of papers per year with the word "personality" in their title (sum over IEEE Xplore and ACM Digital Library).

2 Personality and its Measurement

The key-assumption of personality psychology is that stable individual characteristics result into stable behavioral patterns that people tend to display, at least to a certain extent, independently of the situation. Therefore, the main goals of personality psychology are "*to distinguish internal properties of the person from overt behaviours, and to investigate the causal relationships between them*" [2]. In other words, personality psychology aims at predicting observable individual differences based on stable, possibly measurable, individual characteristics.

Different theories adopt different "*internal properties*" as a personality basis, including physiology (the *biological* perspective), unconscious (the *psychoanalytic* perspective), environment (the *behaviorist* perspective), inner states (the *humanistic* perspective), mind (the *cognitive* perspective), etc. (see [3], [11] for extensive surveys). However, the models that most effectively predict measurable aspects in the life of people are those based on *traits*, a construct widely recognized as "*one of psychology's major achievements*" [12].

Trait models build upon human judgments about semantic similarity and relationships between adjectives that people use to describe themselves and the others [13]. While numerous and widely different, the terms used to describe people typically account for only a few, major dimensions. These latter, if sufficiently stable, are then adopted as *personality traits*, i.e. as factors capable of capturing stable individual characteristics underlying overt behavior. The main criticism against this type of models is that traits are purely descriptive and do not correspond to actual characteristics of individuals [14]. On the other hand, several decades of research and experiments have shown that the same traits appear with surprising regularity across a wide spectrum of situations and cultures, suggesting that they actually correspond to psychologically salient phenomena [12]. These traits, known as *Big-Five* (BF) or *Five-Factor Model* (FFM), are today the "*the dominant paradigm in personality research, and one of the most influential models in all of psychology*" [15].

Trait based models are widely accepted in the computing community as well. All of the works surveyed in this article adopt personality traits (the BF in 76 cases out of 81) and, to the best of our knowledge, no other theories were ever adopted in computing oriented research. On one hand, this barely reflects the dominant position of trait based models in personality psychology. On the other hand, trait models represent personality in terms of numerical values (see below), a form particularly suitable for computer processing.

2.1 The Big-Five and Their Measurement

The Big-Five traits are as follows:

- *Extraversion*: Active, Assertive, Energetic, Outgoing, Talkative, etc.
- *Agreeableness*: Appreciative, Kind, Generous, Forgiving, Sympathetic, Trusting, etc.
- *Conscientiousness*: Efficient, Organized, Planful, Reliable, Responsible, Thorough, etc.
- *Neuroticism*: Anxious, Self-pitying, Tense, Touchy, Unstable, Worrying, etc.
- *Openness*: Artistic, Curious, Imaginative, Insightful, Original, Wide interests, etc.

Attempts were made to enrich the trait set with more dimensions, but "*Five-factor solutions were remarkably stable across studies, whereas more complex solutions were not*" [16]. Other models considered less traits [17], but these still appeared to be linear combinations of the BF. In other words, the BF "*provide a set of highly replicable dimensions that parsimoniously and comprehensively describe most phenotypic individual differences*" [18].

Intuitively, assessing the personality of an individual means to measure how well the adjectives above describe her. Questionnaires where people rate their own behavior with Likert scales are the instrument most commonly adopted for such a purpose [19]. The most popular include the *NEO-Personality-Inventory Revised* (NEO-PI-R, 240 items) [20], the *NEO Five Factor Inventory* (NEO-FFI, 60 items) [21], and the *Big-Five Inventory* (BFI, 44 items) [22] (see [2] for an extensive survey). Short questionnaires (5-10 items), much faster to fill, were built by retaining only those items that best correlate with the results of the full instruments [23], [24] (Table 1 shows the BFI-10, the short version of the BFI).

ID	Question	Trait
1	I am reserved	Ext
2	I am generally trusting	Agr
3	I tend to be lazy	Con
4	I am relaxed, handle stress well	Neu
5	I have few artistic interests	Ope
6	I am outgoing, sociable	Ext
7	I tend to find fault with others	Agr
8	I do a thorough job	Con
9	I get nervous easily	Neu
10	I have an active imagination	Ope

TABLE 1
The BFI-10 [23] is the short version of the Big-Five Inventory. Each Item is associated to a Likert scale (from "*Strongly disagree*" to "*Strongly agree*") and contributes to the score of a particular trait. The answers are mapped into numbers (e.g., from -2 to 2). In general, the questionnaires are available in multiple languages. The translations aim not only at ensuring that the questions are understandable to native speakers of different languages, but also that people from different cultures assign the same meaning to the traits.

A questionnaire is expected to possess good *validity*, i.e. "*[to actually measure] the underlying construct it claims to assess*" [2]. In the case of personality, validity can be analyzed under several points of view, but the one that seems to be more relevant to this work is the *criterion validity*, i.e. "*the extent to which the test correlates with some independent index measured at the same time as the test is administered*" [2]. For example, if an extravert person is expected to have more social contacts than others, then there should be a significant correlation between extraversion (as measured with the questionnaire) and number of connections in social media. The value of such a correlation is called *validity coefficient* and it expresses, on one hand, how strongly the traits predict the criterion and, on the other hand, how valid is the questionnaire.

First person questionnaires like those of Table 1 lead to *self-assessments* and are traditionally considered to yield the *true* personality of an individual [23]. Third person questionnaires (where, e.g., "*I tend to be lazy*" becomes "*this person tends to be lazy*") lead to *attributions* and result into the personality others attribute to a given individual. In the latter case, every subject must be rated by several assessors and each of these must rate all subjects involved in an experiment. Statistical criteria (e.g., the reliability proposed in [25]) allow one to set the number of assessors based on their mutual agreement.

The main limitation of self-assessments is that the subjects might tend to bias the ratings towards socially desirable characteristics, especially when the assessment can have negative consequences like, e.g., failing a job interview. Therefore, an item like "*I tend to be lazy*" might be rated with *disagree* simply because the subjects try to convey a positive impression and hide negative characteristics. However, extensive experiments have shown that the correlation tends to be high between self-assessments and assessments provided by acquainted observers (spouses, family members, etc.) [19]. This proved to be a major step towards the acceptance of questionnaires as a means of personality assessment. Furthermore, it paved the way to the study of personality perception, i.e. the attribution of personality traits to others. While not necessarily corresponding to the true personality of an individual, perceived traits are important because they determine, to a large extent, the behavior that people adopt towards a given individual [5].

2.2 Personality: from Psychology to Computing

Several works investigate the interplay between personality and computing by measuring the link between traits and use of technology [26], [27], [28], [29], [30], [31], [32]. The core principle behind this line of research is that users externalize their personality through the way they use technology. Therefore, personality traits should be predictive of users' behavior.

The study in [26] shows that personality traits, in particular Neuroticism and Openness, predict to a significant extent whether a person activates or not a blog (according to a study done over 278 undergraduate students in the USA) [26]. Other works aim at predicting the effect of personality on observable social media behavior [27], [28], [29]. The first study [27] applies the Linguistic Inquiry Word Count (see Section 4.1) to analyze the Tweets produced over one month by 142 Twitter users that have filled the Big Five Inventory (see Section 2.1). The results show that there is significant correlation between the frequence of several LIWC categories (e.g., articles, auxiliary verbs, affective processes and positive emotions) and the Big Five traits. The only trait that does not show significant correlation with any of the LIWC categories is Conscientiousness. The privacy on Facebook is the focus of [28], where the personality traits of 1323 users - collected with *myPersonality* [33] - are used as features to predict their respective Item Response Theory (IRT) scores, i.e. the psychometric measurements of their attitude towards the privacy problem (*what is private and what is not*). While not leading to statistically significant results, the investigation still shows that personality traits explain in part the tendency to self-disclosure. In the case of [29], the personality traits of 652 subjects are used to predict the motivations behind the use of Youtube. All traits are correlated, to a statistically significant extent, with at least one of the dimensions adopted to represent the motivations behind the use of Internet [34].

The effect of personality on the tendency to use or not mobile phones in certain contexts (e.g., in public spaces where unacquainted individuals can hear the conversation) is the focus of [30]. The work considers 42 individuals and measures their attitude towards incoming calls (or the possibility of making a call) when others

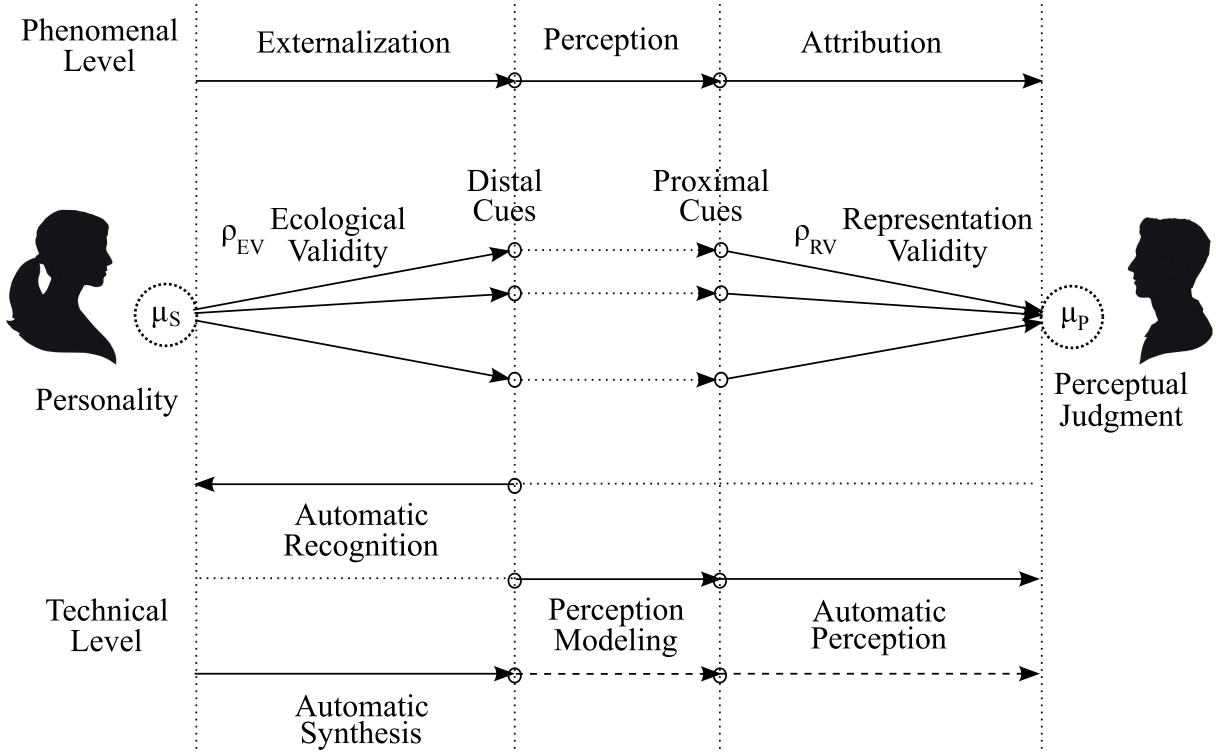

Fig. 2. The Figure shows the relationship between the Brunswik Lens and the three main problems addressed in Personality Computing. Automatic Personality Recognition is the inference of self-assessments (μ_S in the figure) from distal cues, Automatic Personality Perception is the inference of assessments (μ_P in the figure) from proximal cues, Automatic Personality Synthesis is the generation of artificial cues aimed at eliciting the attribution of predefined traits.

are more or less close. The results show that there is a statistically significant difference between subjects falling in the upper or lower half of the observed Eysenck traits scores (Extraversion, Neuroticism and Psychoticism). In [31], the analysis of 112 subjects shows the correlation between the Big Five traits and the amount of time spent in operations like typing SMS, setting ring tones and screen wallpapers, calling others, etc.

Finally, the experiments of [32] measure the correlation between the Big-Five Traits of 214 game players, Genre Preference and *Player Experience of Need Satisfaction* (PENS), a construct which includes five dimensions accounting for presence/immersion, relatedness, intuitive controls, competence and autonomy. The results show weak, but significant correlations between Neuroticism and presence/immersion, Agreeableness and intuitive controls, and Openness and autonomy.

The psychological work presented in this section mirrors the interest of the computing community for the interplay between personality and blogs (see Section 4.1), social media (see Sections 4.3 and 5.3), mobile devices (see Section 4.4), and computer games (see Section 6).

3 Personality Computing

While considering a wide spectrum of scenarios and contexts, Personality Computing approaches appear to address only three fundamental problems, namely *Automatic Personality Recognition* (APR), *Automatic Personality Perception* (APP) and *Automatic Personality Synthesis* (APS). According to [35], the three areas stem from different aspects of the *Brunswik Lens* [36], the cognitive model depicted in Figure 2. Originally proposed to explain how living beings gather information in the environment, the Brunswik Lens was later adopted to describe *externalization* and *attribution* of socially relevant characteristics during human-human [37] and, more recently, human-machine [35] interactions.

The rest of this section describes the Lens Model in detail and shows the correspondence between the phenomena the Lens accounts for and the three Personality Computing problems mentioned above.

3.1 Personality Externalization and APR

According to the Lens Model, individuals *externalize* their personality through *distal cues*, i.e. any form of observable behavior that can be perceived by others (see left hand side of Figure 2). In other words, while being an abstract psychological construct non accessible to direct observation, personality leaves physical traces - or markers - in virtually everything observable individuals do [38].

Automatic Personality Recognition targets the externalization process and it is the *task of inferring self-assessed* *personalities from machine detectable distal cues*. The task is called "*Recognition*" because it aims at inferring traits resulting from self-assessments (see Section 2.1), traditionally considered to be the true traits of an individual [23]. In most cases, APR approaches adopt methodologies typical of Affective Computing, Social Signal Processing, sociolinguistics and the other domains aimed at inferring emotional and social phenomena from machine detectable behavioral evidence. Any measure of the covariation between personality traits and distal cues (typically the correlation or the Spearman Coefficient) is referred to as *Ecological Validity* of the cues. In computing research, covariation studies are often aimed at performing feature selection, i.e. at identifying the distal cues most likely to lead to high APR performance. Section 4 surveys the main APR works presented in the literature.

3.2 Personality Attribution and APP

Distal cues that reach an observer undergo a perception process that results into a *percept*, i.e. the "*mental representation of something that is perceived*" [39] (see central part of Figure 2). For example, a listener does not perceive the energy of the acoustic waves emitted by a speaker, but how loud this latter speaks. For this reason, the Lens Model distinguishes between distal and *proximal* cues, these latter being the ones the observer actually perceives (in the example above, energy and loudness are the distal and proximal cue, respectively). Proximal cues activate the *attribution* process (see right hand side of Figure 2), i.e. the development of a *perceptual judgment* that accounts for the personality traits an observer attributes to a person being observed.

Automatic Personality Perception is the task of inferring the personality observers attribute to a given individual from proximal cues. Unlike the case of APR, the target of APP is not the true personality of individuals, but the personality these are attributed by others. Therefore, while APR relies on self-assessments, APP adopts assessments made by others about the subjects under examination. The methodologies that work for APR are effective for APP as well, but current approaches are unable to use proximal cues and use distal cues as an approximation instead. Measures of the covariation between proximal cues and attributed personality traits are referred to as *representation validity* of the cues. Like in the case of APR, APP works often include covariation studies aimed at identifying the cues most likely to result into high APP performance.

APP approaches typically aim at predicting the average of the traits attributed by multiple raters. These latter differ in terms of status, disposition, personality, etc. and, hence, their assessments are different. The individual assessments are the result of an actual attribution process. Therefore, they can be considered real attributed personalities. The average of the assessments does not result from an attribution process and, hence, cannot be considered a real personality. However, the prediction of the average remains an important task because it captures what is common across individually assigned traits and, furthermore, it can provide indications on the factors driving the attribution process. Section 5 surveys the main APP works presented in the literature.

3.3 Artificial Cues, Personality Attribution and APS

One of the main findings of social cognition is that people spontaneously and unconsciouly assign socially relevant characteristics, including personality traits, to any individual they meet [5]. The phenomenon is so natural and pervasive that it applies not only to people, but also to any device that exhibits human-like features: "*Give anything eyes and a mouth [...] and personality responses follow*" [40].

Automatic Personality Synthesis is the task of automatically generating distal cues aimed at eliciting the attribution of desired personality traits. In terms of the Brunswik Lens (see Figure 2), APS includes both externalization and attribution processes. In the case of the externalization, the cues are not generated by humans, but by any machine capable of displaying human-like behaviors (robots, avatars, artificial agents, etc.). In the case of the attribution, the process involves human observers that assign, typically unconsciously, personality traits to the machine. The main goal of the process is to ensure that the traits assigned by the observers correspond to those planned by the machine designers. Section 6 surveys the main APS works presented in the literature.

4 Automatic Personality Recognition

APR approaches presented so far in the literature consider a wide spectrum of distal cues, including written texts, nonverbal behavior, data collected via mobile or wearable devices and online games. The experiments are based on self-assessments and this makes it possible, in a few cases, to perform experiments over several thousands of subjects.

4.1 Text Based APR

Language psychology shows that the choice of words is driven not only by meaning, but also by psychological phenomena such as emotions, relational attitudes, power status and personality traits: "*Words and language [...] are the very stuff of psychology [...] the very medium by which cognitive, personality, clinical, and social psychologists attempt to understand human beings.*" [41]. Therefore, integrating sociolinguistics in techniques for automatic text analysys makes it possible, among other tasks, to infer personality traits from written texts [42], [43], [44], [45], [46], [47], [48], [49].

One of the earliest efforts in this direction was proposed in [42]. The experiments were performed over 2263 essays written by roughly 1200 students that filled the NEO-FFI (see Section 2.1). However, only subjects in the upper and lower third of observed Extraversion and

Neuroticism scores were retained for the tests. The words were grouped into four psychologically meaningful categories: *function* (articles, prepositions and other words aimed at making sentences grammatically correct), *cohesion* (terms that help to make reference to the context such as deictic expressions and pronouns), *assessment* (terms that evaluate the content in terms of validity, likelihood, desirability, etc.) and *appraisal* (terms that express the attitude of the writer towards the content). The texts were represented with the relative frequencies of the words appearing in each category. The goal of the experiments was to discriminate between subjects at the opposite extremes of Extraversion and Neuroticism (see above). The task was performed with Support Vector Machines and the best reported accuracy was around 58% for both traits.

The same data (with around 250 additional samples) and approach were used in [43], but adopting 88 word categories from *Linguistic Inquiry and Word Count* (LIWC), a psychologically oriented tool for text analysis [41], and the features of the *Medical Research Council* Psycholinguistics Database (MRC), a dictionary including 150837 entries each accompanied by up to 26 attributes. The task was the discrimination between individuals scoring in the upper and lower half of the observed scores for all Big Five traits. The accuracies ranged between 50% and 62% depending on trait and classification approach. The best performance was achieved for Openness using Support Vector Machines. The same work presents similar experiments over the EAR (Electronically Activated Recorder) corpus [50], a collection of random conversation snippets involving 96 subjects. In this case as well, the best accuracy is around 62% for Openness, but using a Naive Bayes classifier.

Most recent efforts aimed at inferring personality traits from texts tend to focus on blogs [46], [47], [48], [44], [45]. The main reason is that these tend to focus on personal issues and experiences, therefore they are likely to show traces of their author's personalities [51].

In [46], [47], the experiments involved 551 subjects assessed in terms of the *Egogram* [52], a personality model based on five communication oriented traits, namely Critical Parent (CP), Nurturing Parent (NP), Adult (A), Free Child (FC), and Adapted Child (AC). The features were the frequencies of the words (grouped into *adjectives*, *adverbs*, *conjunctions*, *exclamations*, *internet slang* and *emoticons*) showing the highest information gain with respect to the scores associated to the traits. The inference was performed with Naive Bayes classifiers using 2, 3 and 5 classes. The F-meaure reached up to 85% depending on number of classes, trait and word category.

In [48], the experiments aimed at predicting whether a blogger was "*low*" (score 0 or 1), "*medium*" (score 2 or 3) or "*high*" (score 4 or 5) with respect to each of the Big Five traits. The task was performed with ordinal logistic models where the LIWC categories were used as independent variables and the classes above as dependent ones. The correlation between actual and predicted classes was low (around 0.1 for all traits), but the analysis of the model parameters (performed over 5042 posts written by 2393 bloggers) allowed the authors to identify some of the main motivations driving the bloggers. For example, neurotic authors tend to blog to release their tensions while extraverts tend to talk about their life.

The works presented so far in this section adopt lexical approaches, i.e. they are based on statistics over the use of individual words. Other approaches consider word N-grams [44], [45] - N-long word sequences - or parse the texts under analysis [49]. The works proposed in [44], [45] avoid the adoption of sociolinguistics and represent texts with the frequency of N-grams in the text (with $N = 2$ and $N = 3$). The goal of the experiments was to discriminate between high and low scoring individuals for all Big Five traits except Openness. In [44], the experiments were performed over the blogs of 71 subjects and the accuracy ranged from 45% (random) to 100% depending on N-gram selection and definition of the classes. In [45], the same experiments were repeated over the blogs of 1672 subjects and the top accuracy decreased to roughly 65%. In both cases, the classification was performed with Naive Bayes classifiers and Support Vector Machines. In the case of [49], the experiments were performed over 145 essays on artificial life written by undergraduate students. The authors apply a text parsing technique to extract N-grams of parts-of-speech (e.g., *subject*, *object*, etc.) and then predict automatically the *Myers-Brigg Type Indicators*, i.e. whether a person belongs to the negative or positive pole along four dimensions: *Attitudes*, *Information-Gathering*, *Decision-Making* and *Lifestyle* [53]. The average F-score ranges between 49.1% (Decision-Making) and 65.4% (Attitude).

4.2 APR and Nonverbal Communication

Psychology suggests that nonverbal communication is, at the same time, an externalization of personality [54] and a cue that influences the traits that others attribute to an individual [5]. From an APR point of view, this means that people's personality can be inferred, at least in principle, from automatically detected nonverbal behavioral cues. Such a key-idea underlies several works that perform APR based on nonverbal aspects of verbal communication (everything in speech except words) [43], [55], interpersonal distances [56] and multimodal combinations of speaking style (prosody, intonation, etc.) and body movements [57], [58], [59], [60], [61] (see Table 3 for a synopsis of data, approaches and results).

In [43] (this work mixes both verbal and nonverbal cues), the tests were performed over the EAR corpus (see Section 4.1) and the features were mean, extremes and standard deviation of pitch, intensity and speaking rate, LIWC and MRC. The experiments aimed at discriminating between individuals in the upper and

Ref.	Subj.	Samples	Features	Task	Ext.	Agr.	Con.	Neu.	Ope.	Other
[42]	> 1200	2263 written essays	word category frequencies	C(2)	58.0 ACC			58.2 ACC		
[43]	> 1200	2479 written essays	LIWC, MRC	C(2)	56.3 ACC	56.3 ACC	55.6 ACC	58.2 ACC	62.1 ACC	
[43]	96	96 conversation transcripts	LIWC, MRC prosody	C(2)	57.3 ACC	58.3 ACC	53.2 ACC	50.4 ACC	61.4 ACC	
[46], [47]	551	551 blog posts	word category frequencies	C(2) C(3) C(5)						F-measure up to 85% on Egogram traits
[48]	2393	5042 blog posts	selection of LIWC categories	OR(3)						Correlation ≈ 0.1 for all Big 5 traits
[44]	71	71 blog posts	N-grams	C(2) C(3) C(5)	100 90.1 44.7 ACC	100 90.4 69.8 ACC	100 92.3 62.0 ACC	96.0 94.4 49.3 ACC		
[45]	1672	1672 blog posts	N-grams	C(2) C(3)	55.4 44.2 ACC	61.6 46.6 ACC	64.8 47.4 ACC	56.3 40.2 ACC		
[49]	145	145 essays	N-grams of parts-of-speech	C(2)						F-measure up to 65.4% on Myers-Brigg Type Ind.

TABLE 2

Text based APR. The table reports, from left to right, the number of subjects involved in the experiments, number and type of behavioral samples, main cues, type of task and performance over different traits. The column "Other" refers to works using models different from the Big-Five. C(n) for Classification with n classes, OR(n) for Ordinal Regression with n classes and ACC for accuracy.

Ref.	Subj.	Samples	Features	Task	Ext.	Agr.	Con.	Neu.	Ope.	Other
[43]	96	96 conversation transcripts	prosody LIWC, MRC	C(2)	57.3 ACC	58.3 ACC	53.2 ACC	50.4 ACC	61.4 ACC	
[55]	12	119 conversations	OpenSMILE speech features	C(2)	63.0 ACC	56.3 ACC	95.0 ACC	32.8 ACC	40.3 ACC	
[56]	13	2 interactive sessions	interpresonal distances walking velocity	C(2)	66.0 ACC			75.0 ACC		
[57]	48	12 meetings of 4 persons	prosody, speech activity body movements	C(3) C(3)	94.4 85.0 ACC					ACC for LOC = 94.9 ACC for LOC = 86.0
[58]	89	89 self presentations	prosody, posture, face/hand/head movements	C(2)	70.8 ACC	65.2 ACC	73.0 ACC	76.4 ACC	66.3 ACC	
[59]	43	4 collaborative tasks per subject	prosody, turn takings motion activity	C(2)	81.4 ACC	69.8 ACC	69.8 ACC	81.3 ACC	60.5 ACC	

TABLE 3

APR and nonverbal communication. The table reports, from left to right, the number of subjects involved in the experiments, number and type of behavioral samples, main cues, type of task and performance over different traits. The column "Other" refers to works using models different from the Big-Five. C(n) for Classification with n classes, LOC for Locus of Control and ACC for accuracy (percentage of correctly classified samples).

lower half of the observed scores of each trait. The best performance, 61% accuracy, was achieved for Openness. However, the performance was at the chance level for all traits. In [55], the experiments were performed over the *PersIA* corpus, a collection of 119 conversations involving 24 subjects simulating the interaction between tourists and tour operators (only the 12 subjects playing the latter role were used for the tests). A total of 6552 features available in openSMILE [62] were extracted and speech samples were classified as "*High*" (above median) or "*Low*" (below median) using *boostexter*. The best results were obtained for Extraversion (63% accuracy) and Conscientiousness (95% accuracy). The approach in [56] makes use of interpersonal distances and walking velocity to predict whether 13 subjects (an ad-hoc collected corpus) are above or below median with respect to Extraversion and Neuroticism. The accuracies are 66% and 75%, respectively.

The approaches in [57] consider one minute long segments extracted from 12 meetings each involving four different subjects (the "*Mission Survival Corpus*" [60]). The features include prosody measurements (mean and standard deviation of formants, spectral entropy, autocorrelation peaks, energy, etc.), speech activity (percentage of speaking time per subject, number and length of voiced segments, etc.) and energy associated to body movements (captured via Motion History Images). The experiments targeted two personality traits, namely Extraversion and Locus of Control (the tendency to attribute the control of events to others or oneself). The task consisted in correctly assigning behavioral samples of one subject to lower, middle or upper segment of observed trait score ranges. The accuracy of different classifiers (Support vector Machines with different kernels) was higher than 90%, especially when using features that account for the context (i.e. what others than the subjects under exam do). The latter aspect was considered in particular in [61], where the attention received by others in terms of gaze was shown to be an effective predictor of Extraversion.

The experiments of [58], were performed over 89 video self-presentations delivered via Skype. Prosodic features (statistics of pitch, intensity and duration of voiced segments), eye-gaze direction, frown (amount and length), posture (duration of back and forward leaning events), hand movements, head shakes/nods, fidgeting and duration of the videos were fed to Support Vector Machines and Naive Bayes Classifiers to predict whether one individual was above or below median with respect to the Big Five Traits. The accuracy ranged between 65% and 76% depending on trait, subset of features adopted and classifier. Similar binary experiments were performed in [59], where behavior in human-machine interactions (prosodic features, turn taking and motion activity) was used to infer whether subjects were above or below median with respect to the Big Five traits. The best results were achieved for Extraversion and Neuroticism, but at different collaboration levels of the interaction settings.

4.3 APR on Social Media

Social media are one of the main channels through which people interact with others, an ideal means for self-disclosure and, therefore, an excellent ground for research on personality computing [51], [63], [64], [65], [66], [67], [68], [69] (see Table 4 for a synopsis of data, approaches and results).

The approach proposed in [51] analyzes Facebook profiles based on whether they post or not certain personal characteristics (e.g., name, education, religion, marital status, etc.), the density of their egocentric networks (percentage of possible links that actually exist between their friends), the amount of characters used to describe favorite activities, the number of organizations users belong to, and whether political orientations are posted or not. Furthermore, the text posted in the profiles (e.g., in the "*About Me*" section) was analyzed with LIWC and several usage statistics (e.g., time since the last update of the profile). Regression experiments were performed over 167 users assessed in terms of the Big Five, and the results show that Gaussian Processes and M5 algorithm allow one to predict personality scores with a mean absolute error lower than 0.13. A similar approach was used by the same authors to predict the personality of Twitter users [63]. In this case, the experiments were performed over the 2000 latest Tweets of 279 users. The features included not only the frequency of LIWC and MRC categories, but also measurements specific of Twitter (e.g., number of followers and following, number of "*hashtags*", etc.). The prediction experiments led to a mean absolute error between actual and predicted traits between 0.11 and 0.18 depending on the traits.

In a similar fashion [64], the M5 algorithm predicts the Big Five traits of 335 Twitter users (Root Mean Square Error between 0.6 and 0.9 depending on the trait) with only three features, namely the number of following, followers and people that include the user in their reading list (all numbers are publicly available). Furthermore, the authors of [65] adopted C4.5 decision trees to predict whether 209 users of "*RenRen*" - the Chinese version of Facebook - are in the upper, middle or lower segment of observed personality scores. The F-measure ranges between 70% (Agreeableness) and 72% (Extraversion). In this work, the features included information on users (e.g., gender and age), usage statistics (e.g., data upload frequence, amount of posts per time unit, etc.) and measurements accounting for the emotional state. The techniques described so far for predicting personality traits were adopted in [66] to predict the personality traits of 156 Italian users on "*FriendFeed*". Features were mainly based on text analysis (exclamation marks, punctuation, self references, word count, etc.) and the evaluation was made by measuring how stable predicted traits were across multiple posts of the same user (average accuracy 63.1%). In this respect, this work is an attempt of recognizing personality traits without previously collecting self-assessments, an approach that might be useful to investigate large populations of users for which it might be difficult to collect questionnaires. In a similar fashion, the approach of [67] labels 10000 users of *Livejournal* (a blogging site) as introverts or extroverts (5000 per category) depending on the number of their friends: $1-3$ for the former and $108-150$ for the latter. LIWC features and logistic regression achieved an F-measure of around 80% in assigning the samples to the correct class (see Section 4.1). The last approach [68] considers 300 Flickr users and the pictures these post as *favorite*. For each user, the approach considers 200 favorite pictures, represents them with a counting grid model (see the paper for more details) and then applies a regression approach to perform personality recognition. The correlation between actual and predicted traits ranges between 0.17 and 0.22 depending on the traits. However, statistically significant results are obtained only for Openness.

The work in [69] presents the results of the "*Workshop on Computational Personality Recognition*" [2], an initiative where participants were required to work over the same data, i.e. a subset of *myPersonality* [33] including 250 Facebook users and 9900 status updates. The participants were left free to decide about their own experimental protocol and performance metric. The analysis of the

2. http://mypersonality.org/wiki/doku.php?id=wcpr13

Ref.	Subj.	Samples	Features	Task	Ext.	Agr.	Con.	Neu.	Ope.	Other
[51]	167	167 Facebook Profiles	profile info., egocentric networks, LIWC	R	0.12 MAE	0.10 MAE	0.10 MAE	0.11 MAE	0.10 MAE	
[63]	279	2000 tweets per subject	LIWC, MRC, profile info.	R	0.16 MAE	0.13 MAE	0.14 MAE	0.18 MAE	0.12 MAE	
[64]	335	335 Twitter profiles	number of followers/ followings, listed counts	R	0.88 RMSE	0.79 RMSE	0.76 RMSE	0.85 RMSE	0.69 RMSE	
[65]	209	209 RenRen profiles	Profile info., usage statistics, emotional states	C(2) C(3)	83.8 71.7 F	69.7 72.3 F	82.4 70.1 F	74.9 71.0 F	81.1 69.5 F	
[66]	156	473 posts on FriendFeed	Some LIWC categories	U						average accuracy 63.1
[67]	10000	10000 blog posts	LIWC	C(2)	80.0 ACC					
[68]	300	$60,000$ favorite pictures	visual patterns, aesthetic preferences	R	0.19 ρ	0.17 ρ	0.22 ρ	0.12 ρ	0.17 ρ	

TABLE 4

APR on social media. The table reports, from left to right, the number of subjects involved in the experiments, number and type of behavioral samples, main cues, type of task and performance over different traits. The column "Other" refers to works using models different from the Big-Five. R stands for regression, U stands for unsupervised, Classification and C(n) for Classification with n classes. The performance for the classification tasks is reported in terms of Mean Absolute Error (MAE), Root Mean Square Error (RMSE), F-Measure (F) and accuracy (ACC).

Ref.	Subj.	Samples	Features	Task	Ext.	Agr.	Con.	Neu.	Ope.	Other
[31]	112	112 mobile phone usage questionnaires	demographic info., mobile phone usage	CA						$0.27 < \|\rho\| < 0.32$
[70]	67	67 sociometric badge recordings	speech/physical activity, interaction, proximity	CA						$0.39 < \|\rho\| < 0.46$
[71], [72]	117	117 smartphone usage logs	logs of SMS/calls/ app./bluetooth, profile	C(2)	77.0 F	77.0 F	78.0 F	75.0 F	74.0 F	
[73]	53	53 mobile-phone usage logs	social network measurements	C(2)	79.7 ACC	73.6 ACC	76.9 ACC	73.7 ACC	77.0 ACC	

TABLE 5

APR via mobile and wearable devices. The table reports, from left to right, the number of subjects involved in the experiments, number and type of behavioral samples, main cues, type of task and performance over different traits. The column "Other" refers to works using models different from the Big-Five. CA stands for Correlational Analysis, C(2) for Binary Classification, F for F-Measure and ACC for accuracy (percentage of samples classified correctly). The correlation between features and personality scores is represented with ρ.

contributions leads to a few major indications: applying selection techniques with ranking algorithms to large feature sets appears to be the most effective strategy. Top-down approaches using lexical resources (e.g, the LIWC) work better than bottom-up ones that are based on words and N-grams. Furthermore, including corpora different from *myPersonality* in the training material seems to be beneficial. An important contribution of the workshop is that the corpora have been made publicly available (see URL in footnote), an important step for a domain lacking established benchmarks (see Section 7.1).

4.4 APR via Mobile and Wearable Devices

Mobile phones, like social media, have penetrated our everyday life as quickly and deeply as only a few other technologies did. In 2005, 15 years after mobile phones appeared in the consumer market, there was one subscription every third person in the world, 82 subscriptions every 100 people in the case of Western Europe, the most "mobile" area of the world [74], [75]. Furthermore, while being conceived to exchange phone calls and SMS, standard mobile phones carry an increasing number of sensors (accelerometers, gyroscopes, proximity, etc.) and can be used as wearable devices to "measure" the life of individuals in naturalistic settings [9]. Such a phenomenon has attracted the attention of computing researchers trying to infer personality traits from data collected via mobile phones and wearable sensors [70], [71], [72], [73] (see Table 4.4 for a synopsis of data, approaches and results).

The experiments in [70] use wearable devices to collect behavioural evidence such as speaking activity (speaking time, voiced time, loudness, etc.), movement (intensity, power, etc.), proximity (time in proximity of others, etc.), face-to-face interactions (number of face-to-face interactions, etc.), and position in the social network resulting from mutual proximity (centrality, betweenness, etc.). The results, obtained over a pool of 67 nurses working in the same hospital, consist of several statistically significant correlations between the cues above and personality traits. (e.g., between speaking activity and Agreeableness). In a similar way, the experiments of [71], [72] consider the phone usage logs of 117 subjects over a period of 17 months and provide not only the correlations between a large number of operations (e.g., use of applications, length of calls and SMS, etc.) and the Big Five traits, but also experiments aimed at predicting whether a person is in the lower or upper half of observed personality scores. The results show an F-measure between 40% and 80% depending on traits and particular elements of the usage logs adopted as features.

The approach proposed in [73] focuses on the possibility of using mobile phones to extract social networks based, e.g., on who calls whom over a certain period of time. The experiments are performed over 53 subjects living in a University residence and monitored via mobile phones for 8 weeks. Social networks are represented with *centrality* (out- and in-degree, betweenness, etc.), *efficiency* (e.g., average of inverse path lenght between nodes), *transitivity* (e.g., number of fully connected node triples) and *triadic* (e.g., the percentage of triads where the nodes are not connected with each other) measures (see [76] for an extensive survey on Social Network Analysis features). The experiments aimed at predicting whether each individual was above or below median with respect to the Big Five traits. The accuracies, obtained with Support Vector Machines, ranged between 65% and 80% depending on traits and features.

4.5 APR and Computer Games

Computer games are a significant source of profit (9.8 billion USD in 2009 [77]) and attract increasingly more attention. Therefore, the literature proposes approaches aimed at inferring personality traits from strategies and options players adopt [78], [79] (see Table 6 for a synopsis of data, approaches and results). The key finding of these works is that gaming behavior actually accounts for personality traits.

The work in [78] analyzes the behavior of 1040 players in *World of Warcraft*, one of the most popular Massive Multi-Player Online Role-Playing Games. The subjects are represented in terms of actions, options, strategies, etc. typical of the game like, e.g., number of days in activity, number of competitors "killed", roles played, etc. The application of linear regression approaches leads to a correlation between actual and predicted Big Five scores ranging from 0.2 (Conscientiousness) to 0.3 (Extraversion and Agreeableness). However, the authors mention that such a performance is over-estimated because they adopt the features they observe to be the most correlated with the personality scores. In the experiments of [79], a game is developed to be used as a personality assessment tool. The tests, performed over 50 subjects, show that playing strategies allow the system to predict correctly the personality of an individual in 77.5% of the cases (in tmers of Myers-Briggs Personality Type Indicator [53]).

5 Automatic Personality Perception

APP approaches focus mainly on nonverbal behavior (in particular in speech) and social media. The number of subjects tends to be lower than in the case of APR because the collection of multiple assessments per subject (necessary in the perception case) limits the number of individuals that can be involved in the experiments.

5.1 APP from Paralanguage

Psychologists have been observing that, at least in certain experimental conditions, "*judgments made from speech alone rather consistently [have] the highest correlation with whole person judgments*", where the word "speech" means here not only what people say, but also paralanguage, i.e. everything accompanies words (prosody, vocalizations, fillers, etc.) [54]. The computing literature seems to follow on this core-idea and the number of APP works based on paralanguage is large compared to those based on other modalities (possibly in combination with paralanguage) [43], [80], [81], [82], [83], [84] (see Table 7 for a synopsis of data, approaches and results). Furthermore, speech based APP was the focus of a recent benchmarking campaign [85], [86], the "*Interspeech 2012 Speaker Trait Challenge*" [87], that has led to the first, rigorous comparison of different approaches over the same data and using the same experimental protocol [88], [89], [90], [91], [92], [93], [94], [95], [96]. A large number of features and machine intelligence approaches have been proposed, but none of them appears to clearly outperform the others (see Table 7 and Figure 3 for a synopsis).

The approach in [43], [80] adopts the same prosody features used, in the same two works, for APR (see Section 4.2 or the details). The experiments were performed over the 96 subjects of the EAR Corpus [50] and the personality assessments were obtained by averaging over the scores individually assigned by 6 independent assessors per sample. The experiments aimed at both predicting the exact personality scores and ranking the subjects according to the predicted scores. In the former case, the best result (obtained for Extraversion and Neuroticism) is a reduction by roughly 15% of the error rate made by an approach returning always the average score observed. In the latter case, the best result (obtained for Extraversion) is an accuracy of around 75% in ranking all possible pairs of individuals. To the best of our knowledge, this is the only work where both APR and APP are performed over the same data. The performance is higher in the APP case and the main reason seems to be that, in this task, machines and assessors access the same information, i.e. the behavior of the subjects under analysis. This does not apply to APR where the subjects can assess their own personalities using information that the machines do not necessarily have at disposition (e.g., their personal history). Even though a rigorous comparison is possible only for the experiments in [43], [80],

Ref.	Subj.	Samples	Features	Task	Ext.	Agr.	Con.	Neu.	Ope.	Other
[32]	214	214 online surveys	PENS, genre preference	CA						not reported
[78]	1040	1040 World of Warcraft profiles	gaming behavior	R	0.30 ρ	0.30 ρ	0.20 ρ	0.21 ρ	0.26 ρ	

TABLE 6

APR and computer games. The table reports, from left to right, the number of subjects involved in the experiments, number and type of behavioral samples, main cues, type of task and performance over different traits. The column "Other" refers to works using models different from the Big-Five. CA stands for Correlational Analysis and R for Regression. The correlation between actual and predicted personality traits is identified with ρ.

Ref.	Subj.	Samples	Features	Task	Ext.	Agr.	Con.	Neu.	Ope.	Other
[43]	96	96 conversation transcripts	prosody LIWC, MRC	C(2)	73.0 ACC	61.3 ACC	67.7 ACC	73.9 ACC	57.0 ACC	
				R	79.9 ER	96.7 ER	82.7 ER	86.7 ER	101.6 ER	
				OR	74.0 ACC	69.0 ACC	67.0 ACC	61.0 ACC	63.0 ACC	
[81]	322	640 speech clips	prosody	C(2)	73.5 ACC	63.1 ACC	72.5 ACC	66.1 ACC	60.1 ACC	
[82]	1	30 acted speech clips	prosody, MFCC, spectral features	C(10)						60.0% accuracy personality type classification
[83]	322	640 speech clips	prosody, voice quality	OR(3)	78.6	65.8	70.8	72.0	63.9	
				OR(4)	76.1	63.6	69.4	70.4	61.3	
				OR(5)	75.0	64.6	68.9	69.9	61.6	
				OR(6)	74.9 ACC	64.1 ACC	68.2 ACC	69.0 ACC	61.3 ACC	
[84]	128	32 meetings	speech activity, prosody N-grams, dialog-act, interaction features	C(2)	74.5 ACC	55.4 ACC	67.6 ACC	68.7 ACC	57.1 ACC	

TABLE 7

APP from speech.The table reports, from left to right, the number of subjects involved in the experiments, number and type of behavioral samples, main cues, type of task and performance over different traits. The column "Other" refers to works using models different from the Big-Five. R stands for Regression, OR(n) stands for Ordinal Regression with n classes, C(n) for classification with n classes, ER for Error Rate and ACC for accuracy.

APP experiments seem to lead, on average, to higher performances.

Approaches similar to those of the works above were presented in [81], [82]. In the first works, the experiments were performed over 640 speech clips (10 seconds long) for a total of 322 subjects (the *SSPNet Speaker Personality Corpus*). The number of assessors per clip was 11 in [81]. The experiments aimed at predicting whether the subjects were above or below median with respect to the Big Five traits. The features were statistics (minimum, maximum, mean and relative entropy of differences between consecutive samples) of pitch, energy, first two formants and length of voiced and unvoiced segments. The accuracies achieved with Logistic Regression and SVM were between 60% and 73.5% depending on the traits (Extraversion and Conscientiousness were the best predicted traits). In [81], the parameters of the Logistic Regression provide indications about the features influencing most the decision of the model. The same data was used in [83], where the prosodic features (pitch, energy, etc.) are extracted from the syllable nuclei and the speech representation is completed with voice quality measures (spectral tilt, jittering, etc.). However, the task of this work was not prediciting the traits, but pairwise ranking of individuals according to their traits (performed with an ordinal regression approach). The results show an accuracy between 60% and 78% depending on the particular trait (the best performance is for Extraversion) and the number of ordinal categories considered.

The experiments of [82] were performed over 30 samples of the same professional speaker acting 10 personality types (as per assessed by 20 raters per sample). The features (1450 in total, including Mel Frequency Cepstral Coefficients, Harmonic-to-Noise-Ratio, Zero-Crossing-Rate, etc.) were fed to an SVM and the accuracy in predicting the right personality type was around 60%. Experiments aimed at predicting whether a subject is above or below median with respect to the Big-Five traits were performed in [84]. The results were obtained over the AMI Meeting Corpus (128 subjects acting in a meeting based scenario). The accuracies range between 50% and 74.5% depending on trait and features including speaking activity (e.g., total amount

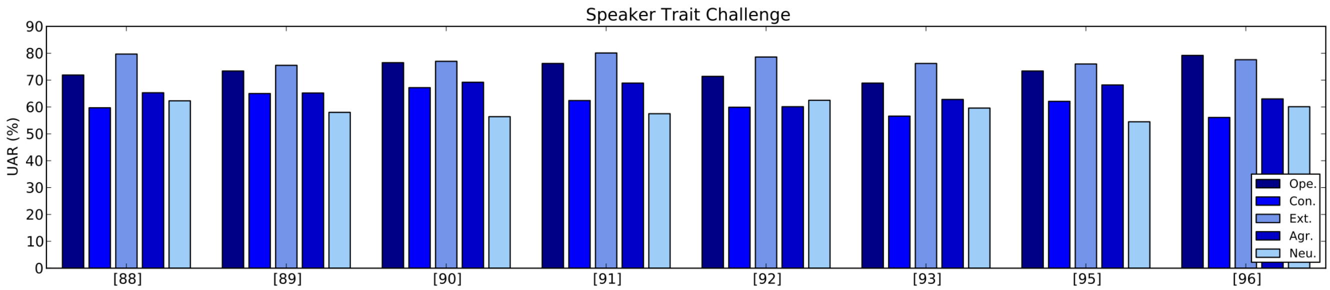

Fig. 3. The chart reports the Unweighted Average Recall of the Speaker Trait Challenge participants. The *SSPNet Speaker Personality Corpus*, used for the experiments, includes 640 speech samples for 322 subjects in total (publicly available at `http://sspnet.eu/2013/10/sspnet-speaker-personality-corpus/`).

of speech for a given subject), prosody (speaking rate, mean, minumum, maximum, standard deviation and median of pitch, etc.), N-gram distributions (see Section 4.1) and Dialogue Acts (e.g., questions, statements, etc.). The combination of all types of features produces statistically significant improvements with respect to the best performing feature type only for Agreeableness and Openness (accuracy around 55% for both traits).

5.1.1 The Interspeech Speaker Trait Challenge

The experiments of the challenge [85] were performed over the *SSPNet Speaker Personality Corpus*. The protocol was the same for all participants and the test data was released - without personality ratings - only at the moment of performing the final experiments. In this way, the chances of over-estimating the performance by adapting or biasing the systems to the test samples were minimized. The task was predicting whether a speech sample is perceived to be above or below median with respect to all traits. The performance measure was the Unweighted Average Recall (UAR). The participants were offered the possibility of using a standard feature set (the 6125 features of openSMILE [62]). See Figure 3 for a synopsis of approaches and results.

The results (see Figure 3) show that no approach clearly outperforms the others. For each trait, the difference between the top approaches is typically not significant. Hence, there seems to be no obvious optimal solution for the prediction of a given trait. Furthermore, the best approach for a certain trait is not necessarily the best approach for the others as well. Therefore, developing trait specific approaches might be a better strategy than developing an approach expected to work on all traits.

Several participants focused on feature selection techniques [88], [89], [90]. The work in [88] starts from the standard feature set of the challenge (see above) and adopts a Set Covering Problem framework to identify the minimum number of features needed to achieve satisfactory performance with Gaussian-Mixture Models. As an alternative, it selects the features with the highest mutual information with respect to the personality traits. The UAR values fall between 59% and 72%. The selection approach used in [89] applies the Sequential Floating Forward Search algorithm to the standard feature set of the challenge. The prediction is then performed with Support Vector Machines and the UAR ranges between 58% and 75% over different traits. In [90], the standard feature set is first enriched with 21760 Modulation Spectrum Analysis features and then submitted to a selection approach that preserves only features that, according to the Kolmogorov-Smirnov test, are distributed in a significantly different way in samples above and below median for a given trait. The prediction is then performed with AdaBoost and the UAR goes up to 77% for Extraversion and Conscientiousness.

Two works emphasize the role of prosody [91], [92]. In the first approach [91], the standard feature set is extended with measurements that account for intonation patterns like, e.g., mean amplitude of peaks, slope of pitch contours, etc. The feature vectors were then fed to Support Vector Machines and the UAR values range roughly between 62% and 80% across different traits. In [92], pitch contours were modeled with polynomials and the parameters of these were used a features, in conjunction with statistics (mean, minimum, maximum, etc.) of pitch and energy, spectral measurements (e.g., the pectral tilt) and Mel Frequency Cepstral Coefficients (MFCC). The prediction was based on Gaussian Mixture Models with Nuisance Compensation and the UAR was between 60% and 76% depending on the trait.

The remaining participants [93], [94], [95], [96] focused on different speech aspects. The approach proposed in [93] adopts an approach, called Anchor Modeling, typically used to index speakers based on their voices. The authors improve the method by performing Within Class Covariance Normalization and then adopt Gaussian Mixture Models to perform the prediction. The speech samples are represented with MFCCs and the UAR fall between 57% and 76% across different traits. The system presented in [94] transcribes automatically the speech samples and then applies LIWC (see Section 4.1) to the resulting texts. Furthermore, it extracts prosodic features like the speaking rate and the duration of pauses, fillers and other nonverbal cues. Trait prediction, performed with Bayesian Networks achieved on average 66% over all traits.

The work in [95] adopts an image processing approach and applies the log-Gabor transform to the spectrogram images. The resulting feature vectors, after applying the Principal Component Analysis for dimensionality reduction purposes, are fed to Support Vector Machines to perform the trait prediction. The performances, expressed in UAR, range between 62% and 76%. In the case of [96], the main contribution consists in automatically dropping the frames (short-term analysis windows) less likely to provide useful information. In particular, the authors cluster the feature vectors they extract from each frame (the standard feature set without estimating statistics) and keep only those vectors that tend to be surrounded by other vectors of the same class (above or below median with respect to each trait). The prediction is performed with Support Vector Machines and the UAR are roughly between 56% and 80%.

5.2 APP and Nonverbal Behavior

Th key-idea of the works presented in this section is that nonverbal behavior can be considered as the physical, machine detectable evidence of social and psychological phenomena [97]. The works presented in Section 5.1 show that this applies to paralanguage in speech (prosody, spectral properties, intonation, etc.). However, nonverbal behavior includes a large number of other cues (facial expressions, gestures, etc.) that, on one hand, are likely to influence the attribution of personality traits and, on the other hand, have already been shown to allow the inference of socially and psychologically relevant information [10]. APP approaches based on nonverbal behavior were proposed in [98], [99], [100], [101] (see Table 8 for a synopsis of data, approaches and results).

The experiments of [98] were performed over a corpus of 442 Youtube video blogs (often called "vlogs") lasting 50 to 70 seconds (each sample shows a different subject). The nonverbal cues included speaking activity (speaking time, length and number of pauses, etc.), prosody (speaking rate, spectral entropy, pitch, etc.), gaze behavior (how much the vlogger looks at the camera, etc.), framing (position of the face in the video frames), motion and combination visual and audio cues (e.g., amount of time spent both looking at the camera and speaking). The goal of the experiments was to predict the personality traits as per assessed by 5 observers. The Root Mean Square Errors range roughly between 0.7 and 1.0 depending on trait and on the feature combination adopted. The approach of [99] proposes to consider "*personality states*", i.e. traits perceived during short episodes and, hence, determined by local rather than global behavioral patterns. The distribution over different personality states across time is then expected to account for the actual traits of an individual. The experiments of [99] were performed over four meeting videos of the Mission Survival Corpus [60], including 16 subjects in total (see Section 4.2). The goal of the experiments was to predict whether each subject falls above or below median with respect to the traits and the accuracies were roughly between 60% and 75%. The features included both speech related measurements (e.g., speaking time, mean energy, pithc, etc.) and social attention (e.g., amount of received and given gaze).

The experiments of [100] adopt 3907 clips extracted from movies where the characters (50 in total) are assessed in terms of the Big-Five traits. The features include not only nonverbal behavior, but also lexical features such as the polarity of the words used. The experiments aimed at predicting the exact personality scores (from 1 to 5) and the accuracies, obtained with Support Vector Machines, ranged from 60% to 85% across different traits. The work in [101] considered a data set of 300 synthetic faces rated along nine different traits (e.g., trustworthy, frightening and extroverted) by 327 judges. The experiments were performed over the faces falling in the upper and lower quartile of each trait. Two approaches were used based on holistic (eigenfaces and Histogram of Gradients) or structural (position of 20 salient points) face representations. The experiments aimed at automatically assigning a test face to the upper or lower quartile of each trait and the accuracy was around 97.8% for the best predicted traits (achieved with GentleBoost, Support Vector Machines, K-Nearest Neighbor and other common algorithms).

5.3 APP and Social Media

Unlike in APR (see Section 4.3), the problem of personality perception over social media has received only limited attention (see Table 9 for a synopsis of data, approaches and results). The approach in [68] uses the images posted as favorite to predict the traits people attribute to 300 Flickr users (see Section 4.3 for more details on data and approach). The correlation between actual and predicted traits range between 0.32 and 0.55 and are, in all cases, statistically significant. The works in [102], [103] investigated the agreement between the actual personality of social media users and the traits they are assigned when they post different types of material. The first work [102], investigates the influence of 440 profile pictures on the traits assigned by 736 unique observers (for a total of 1316 assessments). The pictures were represented in terms of content (e.g., objects, animals, etc.), body portion (if the content is a person), facial expression (type of smile, etc.), appearance (eye glasses, sunglasses, clothes, etc.) and gaze (looking or not at the camera, etc.). The results show the correlation between the measures above and the agreement between actual and assigned traits. The best matching takes place when the profile pictures show an individual, when the person of the pictures smiles, and when the subjects of the picture do not wear hats. The second work [103] does

Ref.	Subj.	Samples	Features	Task	Ext.	Agr.	Con.	Neu.	Ope.	Other
[98]	442	442 Youtube Vlogs	speaking/looking activity, prosody, motion behavior	R	0.80 RMSE	0.87 RMSE	0.74 RMSE	0.79 RMSE	0.67 RMSE	
[99]	16	4 meetings	prosody, speaking activity, social attention	C(2)	73.1 ACC	58.3 ACC	58.3 ACC	63.9 ACC	54.7 ACC	
[100]	50	3907 movie excerpts	prosody, emotion expression, lexical features	C(5)	69.4 ACC	60.3 ACC	81.1 ACC	58.0 ACC	84.8 ACC	
[101]	300	300 synthetic faces	facial appearance, facial salient point	C(2)						see text

TABLE 8

APP and nonverbal behavior. The table reports, from left to right, the number of subjects involved in the experiments, number and type of behavioral samples, main cues, type of task and performance over different traits. The column “Other” refers to works using models different from the Big-Five. R stands for Regression, C(n) for classification with n classes, RMSE for Root Mean Square Error and ACC for Accuracy.

Ref.	Subj.	Samples	Features	Task	Ext.	Agr.	Con.	Neu.	Ope.	Other
[68]	300	60,000 favorite pictures	visual patterns, aesthetic preferences	R	0.55 ρ	0.45 ρ	0.49 ρ	0.54 ρ	0.32 ρ	
[102]	440	440 pictures	photo content, appearance	CA						see text

TABLE 9

APP from social media. The table reports, from left to right, the number of subjects involved in the experiments, number and type of behavioral samples, main cues, type of task and performance over different traits. The column “Other” refers to works using models different from the Big-Five. R stands for Regression, CA for Correlation Analysis and ρ is the Spearman Coefficient.

a similar analysis, but it considers all elements that can appear in an online profile (religion, relationship status, links to pages or videos, etc.). The experiments were performed over 5216 profiles and the results show that the chances of agreement between observers and profile owners are higher when these latter post indications about their beliefs and spirituality, links to funny videos, major sources of satisfaction and joy, etc.

6 Automatic Personality Synthesis

Research in Human-Machine Interaction investigated the role of personality perception in the case of synthetic speech [6], [104], [105], [106], [107], artificial agents [108], [109], [110], [111], [112], [113] and robots [114], [115], [116], [117], [118], [119], [120], [121].

6.1 Speech Based APS

The experiments of [6], [104] show that it is possible to synthesize speech that human listeners tend to perceive as higher or lower in *Extraversion*. The former is obtained by using high volume, pitch of 140 Hz, frequency range of 40 Hz and speaking rate of 216 words per minute, the latter is obtained by setting a volume 15% lower than the extrovert voice, pitch at 84 Hz, frequency rate of 16 Hz and speaking rate of 184 words per minute. Furthermore, the experiments show that human listeners high in *Extraversion* are more likely to accept book recommendations made with extravert-like speech and vice-versa (72 subjects corresponding to the upper and lower *Extraversion* quartile in a group of 150 students). In other words, the subjects seem to favor voices that sound more similar in terms of personality, exactly as it happens in human-human interactions. The manipulation of prosodic parameters in synthetic voices (pitch, pitch range, intensity and speaking rate) was shown to influence the perception of personality in [105], [106]. In the experiments of these works, 36 raters assessed speech samples in terms of how well they fitted the description typically associated to high or level of the Big Five traits. The results show that it is actually possible to stimulate the perception of desired personality traits. The evaluation in [107], probably the earliest attempt of its kind, shows that 50 assessors (25 expert and 25 naïve raters) consistently assign the same traits (practical, intelligent,courteous, etc.) to spoken messages adopted in voice mail systems. Furthermore, the raters identified some traits (e.g., efficient and imaginative) as more desirable than others.

6.2 APS and Artificial Agents

The experiments in [108], [109] show that Embodied Conversational Agents displaying backchannel responses typical of a certain personality traits influence the perception of human observers towards those same traits. In particular, the tests performed in [109] with 187 raters show that differences in physical appearance, activation (body movements, gestures, etc.), face behaviour (more or less direct gaze, facial expressions, etc.) and paralanguage (length and frequency of pauses, hesitations, etc.) result into systematic differences in the attribution of personality traits. In a similar way, the study in [110] shows that combining in all possible ways 2 gesture rates, 4 gesture performance levels (timing, posture and amplitude) and 4 utterances (32 artificial agents in total) influences the perception of *Extraversion*. In particular, the perception of such a trait seems to be heavily influenced by the words of the utterance (40 raters in total). The same approach based on the combination of cues is followed in [111]. All associations of 2 gaze patterns (direct and non-direct), 2 prosody styles (fast, loud and high-pitched versus slow, soft and low-pitched) and 2 eyebrows movements (moving versus not-moving) were assessed in terms of *Extraversion* (24 raters). All cues were found to influence significantly the perception of such a trait. The experiments of [112] propose a model that changes the behavior of a talking head based on personality, mood and emotional state, with the personality modulating the intensity of the agent responses (facial expressions and speech). The results consists of different behavioral displays based on different agent personalities. The effect of head orientation on personality perception is the focus of [113]. The experiments (133 subjects assessing 54 static head poses of the same agent) show that the cue influences the perception of *Extraversion*, *Agreeableness* and *Neuroticism*.

6.3 APS and Robots

The *similarity-attraction* phenomenon investigated for personality coloured speech [6], [104] was studied in the case of robots as well [114], [115], [116], [117], [118], [119], [120]. The experiments of [114], [115] involve 19 subjects undergoing physical rehabilitation therapies. The results show that the introvert subjects (7 in total) tend to spend more time with robots showing low *Extraversion* while the 12 extravert ones have the opposite tendency. The simulation of the trait relied on three main cues: prosody (pitch and volume), verbal content (choice of words and sentences) and proxemics (distance from the subjects). Similar findings result from the experiments in [116], where *Extraversion* is generated via facial expressions. In a pool of 40 subjects, extravert people tend to like "extravert" robots more than "introvert" ones and vice-versa. Facial expressions and color (red or pale) were used in [117] to express six different personalities (high/low *Extraversion*, high/low *Agreeableness*, and high/low *Neuroticism*), but no results on human perception were presented.

In the case of [118], the number of subjects is 28 and the interaction with the robots takes place in a living room. The personality is simulated through space negotiation (straight versus circular trajectories, distance with respect to the subjects), "gaze" behavior (the camera mounted on the robot follows the subjecs or remains static), speaking activity (waiting or not for the subject before talking) and lexical choices. Unlike the other works, the results do not suggest a *similarity-attraction* effect. The same consideration applies to [119], showing that there is no matching between the personality of subjects interacting with a robot and the interpersonal distance adopted during the interaction. However, all subjects tend to get closer to the robot than to a human in the same circumstances. In the experiments of [120], where 48 subjects (24 extravert and 24 introvert) consistently recognized a robot as high or low in *Extraversion* depending on its behavior, but showed a *complementarity-attraction* effect, i.e. introvert subjects preferred extravert robots and vice-versa (personality was simulated using the same prosodic characteristics as those used in [6], [104], led lights, trajectory with respect to the subject, speed and amount of movement). The experiments in [121] involve 31 subjects and try to estimate the influence of culture and context on the perception of a robot's personality. However, the results do not show any significant effect.

7 Towards a Research Agenda

While personality attracts increasingly more interest in the computing community (see Figure 1), a large number of issues and challenges remain still open [35]. This section outlines some of the most important problems that, if correctly addressed, can lead to substantial improvements of the state-of-the-art (the list is not exhaustive).

7.1 The Data

Modern personality psychology starts with the application of statistical, data driven approaches to the words people use to describe others (see Section 2.1) [18], [16]. Not surprisingly, data plays a crucial role in Personality Computing as well and, overall, the lack of widely accepted benchmarks is one of the main limitations of the state-of-the-art (most works propose experiments on data collected *ad-hoc*). To the best of our knowledge, the main publicly available corpora are the *SSPNet Speaker Personality Corpus* [81] - adopted in the *Speaker Trait Challenge* [85] - and the data distributed via *MyPersonality* [33], including the benchmarks used for the *Workshop on Computational Personality Recognition* [69]. Collection and diffusion of standard benchmarks will help to improve the overall level of the domain by allowing rigorous comparisons between different works.

In the case of APP, the main bottleneck is that multiple assessors have to rate all subjects included in the data to make the assessments coherent and consistent (see [25] for the methodological issues related to judgmental studies). This typically limits the size of the APP corpora to a few hundreds of subjects (see tables in Section 3.2). Crowdsourcing [122] might be a potential solution, but it is still unclear whether it can be considered rigorous from a psychological point of view.

In general, the personality assessments included in the data are performed with questionnaires proposed in the literature. These are likely to possess good validity, i.e. to actually measure the personality of individuals. However, explicitly measuring the validity of the questionnaires adopted to build a corpus can make the data collection process more rigorous (see Section 2.1 for more details).

7.2 Methodological Issues

The bulk of the work in personality computing proposes (most often binary) classification and regression approaches mapping behavioral distal cues into personality traits (see Sections 3.2 and 3.1), but it is not clear whether these are the right problems to address. In particular, binary classification approaches - the majority of the works - split subjects into classes (e.g., above and below median with respect to a certain trait) that are not meaningful from a psychological point of view. According to the psychological literature"[...] *a compelling argument can be made for emphasizing comparisons among individuals, which we do in everyday life [...] and which is useful for practical purposes*" [14]. Therefore, ranking people according to their personality traits might be a more suitable and psychologically meaningful task like, e.g., in [43], [48], [83]. Furthermore, the outcomes of APR and APP approaches might be accepted only when satisfying appropriate confidence criteria (e.g., the work in [82] assigns people only the trait that differs most from the mean or median).

So far, personality computing approaches address traits separately because these are supposed to be, by construction, uncorrelated and independent [18]. However, raters' cognitive and cultural biases can determine relationships between traits [123], and the same can happen if the subjects of a given corpus tend to group into categories [124]. In these cases, statistical approaches capable of jointly modeling the traits might have higher performance. However, both the identification of such an approach and the benefits that can result for APP and APR performance are, to the best of our knowledge, an open issue.

A last problem is that the perception process in the Brunswik Lens (see Figure 2) was largely, if not at all, neglected in the personality computing literature. All of the works surveyed in this article adopt distal cues as a basis for their inference techniques. While this is appropriate in the case of APR, it can limit the effectiveness in APP and APS where the goal is to predict and manipulate, respectively, the traits people attribute to others. The problem might be addressed by mapping distal cues - the features extracted from the data at disposition - into percepts, i.e. into conceptual / semantic representations of what people perceive. For example, image pixels represented in terms of Hue, Saturation and Value might be represented in terms of color categories such as red, blue or yellow [68]). As a first approximation, psychoacoustic speech processing techniques [125] or aesthetic oriented image analysis repesentations [126], [127] (the list is not exhaustive) might better account for how people perceive distal cues.

7.3 Applications

The interest for personality in computing is still at a relatively early stage (see Figure 1) and most of the efforts were dedicated so far to establishing the domain, collecting data, developing methodologies and identifying relevant tasks. Still, some early applications of personality computing were presented in the literature. In [128], [129], users' personality ratings improve the performance of a recommender system, the personality of a synthetic voice was shown to increase the acceptance of GPS systems in [104] and the earliest personality coloured speech synthesizers are available on the market [130].

However, Personality Computing is likely to attract attention in a large number of other application domains (the list is not exhaustive). In a context where personal data is considered "*the new oil of the internet and the new currency of the digital world*" [131], personality computing will help to mine the large amount of digital traces people leave online and to make sense of social media users [33], to target advertisement campaigns to the right potential customers [7] or to tune retrieval technologies to users' personality [68].

Following the progress of technologies dealing with autism spectrum desorders and other developmental problems [132], Personality Computing is likely to play a major role in technologies aimed at detecting diseases like paranoia and schizophrenia that typically interfere with personality [133]. In this respect, computer applications (e.g., games) might work as an assessment and capture personality related evidence [78], [79]. Furthermore, assistive technologies involving the use of robots or other types of artificial companions will benefit from synthetic personalities that will increase their acceptance, especially with people that are not familiar with technology [119].

Human Computer Interaction can adopt personality computing not only to sense users and, based on their traits, make informed guesses about their needs and preferences, but also by synthesizing personality traits appropriate for the particular application (e.g., highly conscientious for an artificial tutoring system or highly agreeable for the interface of a counseling service). In more general terms, every application expected to seamlessly integrate our everyday life [97] or interact with humans like humans do [10] is likely to benefit from Personality Computing.

The development of personality computing approaches can be beneficial to personality and social psychology as well. In particular, computing technologies allow the processing of large amounts of behavioral data that might be difficult to analyze with techniques traditionally applied in psychology (observational methods, surveys, etc.). In this respect, personality computing might help to establish links between traits and behavior with an effectiveness that was not possible so far. Furthermore, the works presented in Section 2.2 show that there is a link between personality traits and the use of some of the most popular computing technologies, including blogs, videogames, etc. Therefore, personality psychology can become an important tool for designing new applications, predicting the success of a new product, introducing new features to existing systems or developing user adaptation approaches.

8 CONCLUSIONS

To the best of our knowledge, this paper is the first survey of Personality Computing, the domain aimed at automatic recognition, perception and synthesis of personality.

Current approaches build upon the extensive experience accumulated in computing domains aimed at inferring social and psychological phenomena from machine detectable cues, including Affective Computing (the domain dealing with emotions), Social Signal Processing (the area dealing with nonverbal communication in social interactions), sociolinguistics (the field that analyzes traces of social phenomena in language), etc. However, the state-of-the-art is still fragmented: with a few exceptions, the experiments are performed over ad-hoc, proprietary data. This makes it difficult to identify methodologies or approaches that work better than the others. Furthermore, it is not always clear whether personality assessments are collected with rigorous psychological methodologies, especially when it comes to validity issues (see Section 2.1). On the other hand, personality attracts increasingly more attention in a wide and diverse range of communities (social media, robotics, Human-Computer Interaction, speech processing, etc.). In this respect, the field promises to become a common ground for disciplines and areas that hardly communicated with each other so far.

The computing community clearly privileges trait based models and, in particular, the Big Five. Besides being the dominant paradigm in personality psychology, trait models are particularly suitable for computer processing because they represent personality in terms of continuous numerical scores. Still, most APR and APP works split continuous scores into two classes (e.g., above and below average) and adopt binary approaches. This is probably the most important limitation of the current state-of-the-art, mostly due to the inherent difficulty and ambiguity of the APR and APP tasks.

Attempts to overcome such a limitation can take at least two directions: On one hand, it is possible to focus on the technologies and, for example, improve machine learning techniques to better model the relationship between cues (distal or proximal) and personality traits. On the other hand, it might be useful to look for a tighter integration between computing and human sciences, i.e. to design personality computing approaches according to psychological and cognitive mechanisms related to personality externalization and attribution. Moreover, the long-term goal of Personality Computing is to use traits in order to improve an application or a process. Therefore, the effectiveness from such a point of view might be adopted as a criterion to drive further progress.

The organization of two international benchmarking campaigns in the last couple of years [69], [85] confirms the interest for Personality Computing while consolidating the state-of-the-art at least for APP and APR. The use of common corpora and standard experimental protocols allows the rigorous comparison of different works, an important step forward with respect to the initial situation where every work proposes ad-hoc, proprietary data.

Dealing with humans is one of the most important challenges for computing, whether humans are users, subjects appearing in data to be analyzed, or digital material producers and consumers. Personality, as a construct capable of capturing the salient aspects of an individual, might provide a key to better bridge the gap between people and machines.

ACKNOWLEDGMENTS

The research that has led to this work has been supported in part by the European Communitys Seventh Framework Programme (FP7/2007-2013), under grant agreement no. 231287 (SSPNet), in part by the Swiss National Science Foundation via the National Centre of Competence in Research IM2 (Information Multimodal Information management) and Indo-Swiss Joint Research Project CCPP (Cross-Cultural Personality Perception).

REFERENCES

[1] Theophrastus, *The Characters*, 4th Century BC.
[2] G. Matthews, I. Deary, and M. Whiteman, *Personality Traits*. Cambridge University Press, 2009.
[3] D. Funder, “Personality,” *Annual Reviews of Psychology*, vol. 52, pp. 197–221, 2001.
[4] D. Ozer and V. Benet-Martinez, “Personality and the prediction of consequential outcomes,” *Annual Reviews of Psychology*, vol. 57, pp. 401–421, 2006.
[5] J. S. Uleman, S. A. Saribay, and C. M. Gonzalez, “Spontaneous inferences, implicit impressions, and implicit theories,” *Annual Reviews of Psychology*, vol. 59, pp. 329–360, 2008.
[6] C. Nass and K. Min Lee, “Does computer-synthesized speech manifest personality? experimental tests of recognition, similarity-attraction and consistency-attraction,” *in Journal of Experimental Psychology: Applied*, vol. 7, no. 3, pp. 171–181, 2001.
[7] A. Kaplan and M. Haenlein, “Users of the world, unite! the challenges and opportunities of social media,” *Business Horizons*, vol. 53, no. 1, pp. 59–68, 2010.
[8] L. Rainie and B. Wellman, *Networked: The new social operating system*. The MIT Press, 2012.
[9] M. Raento, A. Oulasvirta, and N. Eagle, “Smartphones an emerging tool for social scientists,” *Sociological Methods & Research*, vol. 37, no. 3, pp. 426–454, 2009.
[10] A. Vinciarelli, M. Pantic, D. Heylen, C. Pelachaud, I. Poggi, F. D’Errico, and M. Schroeder, “Bridging the Gap Between Social Animal and Unsocial Machine: A Survey of Social Signal Processing,” *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 69–87, 2012.
[11] P. Corr and G. Matthews, Eds., *The Cambridge handbook of personality psychology*. Cambridge University Press, 2009.
[12] I. Deary, “The trait approach to personality,” in *The Cambridge handbook of personality psychology*, P. Corr and G. Matthews, Eds. Cambridge University Press, 2009, pp. 89–109.
[13] L. Goldberg, “The structure of phenotypic personality traits,” *American Psychologist*, vol. 48, no. 1, pp. 26–34, 1993.
[14] S. Cloninger, “Conceptual issues in personality theory,” in *The Cambridge handbook of personality psychology*, P. Corr and G. Matthews, Eds. Cambridge University Press, 2009, pp. 3–26.

[15] R. McCrae, "The Five-Factor Model of personality," in *The Cambridge handbook of personality psychology*, P. Corr and G. Matthews, Eds. Cambridge University Press, 2009, pp. 148–161.

[16] J. Digman, "The curious history of the Five-Factor Model," in *The Five-Factor Model of Personality*, J. Wiggins, Ed. Guilford Press, 1996, pp. 1–20.

[17] H. Eysenck, "Dimensions of personality: 16, 5 or 3? - criteria for a taxonomic paradigm," *Personality and Individual Differences*, vol. 12, no. 8, pp. 773–790, 1991.

[18] G. Saucier and L. Goldberg, "The language of personality: Lexical perspectives on the five-factor model," in *The Five-Factor Model of Personality*, J. Wiggins, Ed. Guilford Press, 1996, pp. 21–50.

[19] G. Boyle and E. Helmes, "Methods of personality assessment," in *The Cambridge handbook of personality psychology*, P. Corr and G. Matthews, Eds. Cambridge University Press, 2009, pp. 110–126.

[20] P. Costa and R. McCrae, "Domains and facets: Hierarchical personality assessment using the revised NEO Personality Inventory," *Journal of Personality Assessment*, vol. 64, no. 1, pp. 21–50, 1995.

[21] R. McCrae and P. Costa, "A contemplated revision of the NEO Five-Factor Inventory," *Personality and Individual Differences*, vol. 36, no. 3, pp. 587–596, 2004.

[22] O. John, E. Donahue, and R. Kentle, "The Big Five Inventory - versions 4a and 54," University of California, Berkeley, Institute of Personality and Social Research, Tech. Rep., 1991.

[23] B. Rammstedt and O. John, "Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German," *Journal of Research in Personality*, vol. 41, no. 1, pp. 203–212, 2007.

[24] S. Gosling, P. Rentfrow, and W. Swann, "A very brief measure of the Big-Five personality domains," *Journal of Research in Personality*, vol. 37, no. 6, pp. 504–528, 2003.

[25] R. Rosenthal, "Conducting judgment studies: Some methodological issues," in *The new handbook of methods in nonverbal behavior research*, J. Harrigan, R. Rosenthal, and K. Scherer, Eds. Oxford University Press, 2005, pp. 199–234.

[26] R. Guadagno, B. Okdie, and C. Eno, "Who blogs? personality predictors of blogging," *Computers in Human Behavior*, vol. 24, no. 5, pp. 1993–2004, 2008.

[27] L. Qiu, H. Lin, J. Ramsay, and F. Yang, "You are what you tweet: Personality expression and perception on twitter," *Journal of Research in Personality*, vol. 46, no. 6, pp. 710–718, 2012.

[28] D. Quercia, D. Las Casas, J. Pesce, D. Stillwell, M. Kosinski, V. Almeida, and J. Crowcroft, "Facebook and privacy: The balancing act of personality, gender, and relationship currency," in *Proceeedings of the AAAI International Conference on Weblogs and Social Media*, 2012, pp. 306–313.

[29] T. Yeo, "Modeling personality influences on YouTube usage," in *Proceedings of AAAI International Conference on Weblogs and Social Media*, 2010, pp. 367–370.

[30] S. Love and J. Kewley, "Does personality affect peoples' attitude towards mobile phone use in public places?" in *Mobile Communications*, ser. Computer Supported Cooperative Work. Springer, 2005, no. 51, pp. 273–284.

[31] S. Butt and J. G. Phillips, "Personality and self reported mobile phone use," *Computers in Human Behavior*, vol. 24, no. 2, pp. 346–360, 2008.

[32] D. Johnson and J. Gardner, "Personality, motivation and video games," in *Proceedings of the Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, 2010, pp. 276–279.

[33] M. Kosinski and D. Stillwell. myPersonality research wiki. [Online]. Available: http://mypersonality.org/wiki

[34] Z. Papacharissi and A. Rubin, "Predictors of Internet use," *Journal of Broadcasting and Electronic Media*, vol. 44, no. 2, pp. 175–196, 2000.

[35] F. Pianesi, "Searching for personality," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 146–158, 2013.

[36] E. Brunswik, *Perception and the representative design of psychological experiments*. University of California Press, 1956.

[37] K. R. Scherer, "Personality inference from voice quality: The loud voice of extroversion," *European Journal of Social Psychology*, vol. 8, pp. 467–487, 1978.

[38] K. Scherer, "Personality markers in speech," in *Social markers in speech*. Cambridge University Press, Cambridge, 1979, pp. 147–209.

[39] A. Colman, *Oxford Dictionary of Psychology*. Oxford University Press, 2009.

[40] B. Reeves and C. Nass, *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press, 1996.

[41] Y. Tausczik and J. Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods," *Journal of Language and Social Psychology*, vol. 29, no. 1, pp. 24–54, 2010.

[42] S. Argamon, S. Dhawle, M. Koppel, and J. Pennbaker, "Lexical predictors of personality type," *In Proceedings of Interface and the Classification Society of North America*, 2005.

[43] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, "Using linguistic cues for the automatic recognition of personality in conversation and text," *Journal of Artificial Intelligence Research*, vol. 30, pp. 457–500, 2007.

[44] J. Oberlander and S. Nowson, "Whose thumb is it anyway? classifying author personality from weblog text," in *In Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2006, pp. 627–634.

[45] S. Nowson and J. Oberlander, "Identifying more bloggers: Towards large-scale personality classification of personal weblogs," in *Proceedings of the International Conference on Weblogs and Social Media*, 2007.

[46] A. Minamikawa and H. Yokoyama, "Blog tells what kind of personality you have : Egogram estimation from Japanese weblog," in *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 2011, pp. 217–220.

[47] ——, "Personality estimation based on weblog text classification," in *Modern Approaches in Applied Intelligence*, K. Mehrotra, K. Chilukuri, J. Oh, P. Varshney, and A. Moonis, Eds. Springer, 2011, pp. 89–97.

[48] A. J. Gill, S. Nowson, and J. Oberlander, "What are they blogging about? personality, topic and motivation in blogs," in *Proceedings of the The International AAAI Conference on Weblogs and Social Media*, 2009, pp. 18–25.

[49] K. Luyckx and W. Daelemans, "Using syntactic features to predict author personality from text," in *Proceeedings of Digital Humanities*, 2008, pp. 146–149.

[50] M. Mehl, S. Gosling, and J. Pennebaker, "Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life." *Journal of Personality and Social Psychology*, vol. 90, no. 5, pp. 862–877, 2006.

[51] J. Golbeck, C. Robles, and K. Turner, "Predicting Personality with Social Media," in *Proceedings of the Extended Abstracts on Human Factors in Computing Systems*, 2011, pp. 253–262.

[52] J. Dusay, "Egograms and the constancy hypothesis," *Transactional Analysis Journal*, vol. 2, no. 3, pp. 37–41, 1972.

[53] K. Briggs, *Myers-Briggs type indicator*. Consulting Psychologists Press, 1976.

[54] P. Ekman, W. Friesen, M. O'Sullivan, and K. Scherer, "Relative importance of face, body, and speech in judgments of personality and affect," *Journal of Personality and Social Psychology*, vol. 38, no. 2, pp. 270–277, 1980.

[55] A. V. Ivanov, G. Riccardi, A. J. Sporka, and J. Franc, "Recognition of personality traits from human spoken conversations," in *Proceedings of Interspeech*, 2011, pp. 1549–1552.

[56] G. Zen, B. Lepri, E. Ricci, and O. Lanz, "Space speaks: towards socially and personality aware visual surveillance," in *Proceedings of the ACM International Workshop on Multimodal Pervasive Video Analysis*, 2010, pp. 37–42.

[57] F. Pianesi, N. Mana, A. Cappelletti, B. Lepri, and M. Zancanaro, "Multimodal Recognition of Personality Traits in Social Interactions," in *Proceedings of the International Conference on Multimodal Interfaces*, 2008, pp. 53–60.

[58] L. M. Batrinca, N. Mana, B. Lepri, F. Pianesi, and N. Sebe, "Please, Tell Me About Yourself : Automatic Personality Assessment Using Short Self-Presentations," in *Proceedings of the International Conference on Multimodal Interfaces*, Alicante, 2011, pp. 255–262.

[59] L. M. Batrinca, B. Lepri, N. Mana, and F. Pianesi, "Multimodal Recognition of Personality Traits in Human-Computer Collaborative Tasks," in *Proceedings of International Conference on Multimodal Interaction*, 2012, pp. 39–46.

[60] F. Pianesi, M. Zancanaro, B. Lepri, and A. Cappelletti, "A multimodal annotated corpus of consensus decision making meetings," *Language Resources and Evaluation*, vol. 41, no. 3, pp. 409–429, 2007.

[61] B. Lepri, R. Subramanian, K. Kalimeri, J. Staiano, Pianes, and N. Sebe, "Connecting meeting behavior with extraversion - a systematic study," *IEEE Transactions on Affective Computing*, vol. 3, no. 4, pp. 443–455, 2012.

[62] F. Eyben, M. Woellmer, and B. Schueller, "OpenSMILE: the Munich versatile and fast open-source audio feature extractor," in *Proceedings of ACM International Conference on Multimedia*, 2010, pp. 1459–1462.

[63] J. Golbeck, C. Robles, M. Edmondson, and K. Turner, "Predicting personality from twitter," in *Proceedings of IEEE International Conference on Social Computing*, 2011, pp. 149–156.

[64] D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft, "Our Twitter profiles, our selves: Predicting personality with Twitter," in *Proceedings of the IEEE International Conference on Social Computing*, 2011, pp. 180–185.

[65] S. Bai, T. Zhu, and L. Cheng, "Big-five personality prediction based on user behaviors at social network sites," Cornell University, Tech. Rep., 2012.

[66] F. Celli, "Unsupervised personality recognition for social network sites," in *Proceeedings of the International Conference on Digital Society*, 2012, pp. 59–62.

[67] T. Nguyen, D. Phung, B. Adams, and S. Venkatesh, "Towards discovery of influence and personality traits through social link prediction," in *Proceedings of the International AAAI Conference on Weblogs and Social Media*, 2011, pp. 566–569.

[68] M. Cristani, A. Vinciarelli, C. Segalin, and A. Perina, "Unveiling the multimedia unconscious: Implicit cognitive processes and multimedia content analysis," in *Proceedings of the ACM International Conference on Multimedia*, 2013.

[69] F. Celli, F. Pianesi, D. Stillwell, and M. Kosinski, "Workshop on Computational Personality Recognition (shared task)," in *Proceedings of the Workshop on Computational Personality Recognition*, 2013.

[70] D. Olguin-Olguin, P. Gloor, and A. Pentland, "Capturing individual and group behavior with wearable sensors," in *Proceedings of AAAI Spring Symposium on Human Behavior Modeling*, 2009.

[71] G. Chittaranjan, J. Blom, and D. Gatica-Perez, "Who's who with big-five: Analyzing and classifying personality traits with smartphones," in *IEEE Annual International Symposium on Wearable Computers*, 2011, pp. 29–36.

[72] G. Chittaranjan, J. Bloom, and D. Gatica-Perez, "Mining large-scale smartphone data for personality status," *Personal and Ubiquitous Computing*, 2011.

[73] J. Staiano, F. Pianesi, B. Lepri, N. Sebe, N. Aharony, and A. Pentland, "Friends don't lie - Inferring personality traits from social network structure," in *Proceedings of the ACM International Conference on Ubiquitous Computing*, 2012, pp. 321–330.

[74] K. Kalba, "The global adoption and diffusion of mobile phones," Center for Information Policy Research Harvard University, Tech. Rep., 2008.

[75] ITU, "The world in 2010: ICT facts and figures," International Telecommunication Union, Tech. Rep.

[76] S. Wasserman and K. Faust, *Social Network Analysis: Methods and applications*. Cambridge University Press, 1994.

[77] L. Chen, H. Tu, and E. Wang, "Personality traits and life satisfaction among online game players," *CyberPsychology & Behavior*, vol. 11, no. 2, pp. 145–149, 2008.

[78] N. Yee, N. Ducheneaut, L. Nelson, and P. Likarish, "Introverted Elves & Conscientious Gnomes : The Expression of Personality in World of Warcraft," in *Proceedings of the Annual Conference on Human Factors in Computing Systems*, Vancouver, 2011, pp. 753–762.

[79] C. Yaakub, N. Sulaiman, and C. Kim, "A study on personality identification using game based theory," in *Proceedings of the International Conference on Computer Technology and Development*, 2010, pp. 732–734.

[80] F. Mairesse and M. Walker, "Words mark the nerds: Computational models of personality recognition through language," in *In Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 2006, pp. 543–548.

[81] G. Mohammadi and A. Vinciarelli, "Automatic personality perception: Prediction of trait attribution based on prosodic features," *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 273–278, 2012.

[82] T. Polzehl, S. Moller, and F. Metze, "Automatically assessing personality from speech," *In Proceedings of IEEE International Conference on Semantic Computing*, pp. 134–140, 2010.

[83] G. Mohammadi, A. Origlia, M. Filippone, and A.Vinciarelli, "From speech to personality: Mapping voice quality and intonation into personality differences," in *Proceedings of ACM International Conference on Multimedia*, 2012, pp. 789–792.

[84] F. Valente, S. Kim, and P. Motlicek, "Annotation and recognition of personality traits in spoken conversations from the ami meetings corpus," in *Proceedings of Interspeech*, 2012.

[85] B. Schuller, S. Steidl, A. Batliner, E. Noth, A. Vinciarelli, F. Burkhardt, R. V. Son, F. Weninger, F. Eyben, T. Bocklet, G. Mohammadi, and B. Weiss, "The INTERSPEECH 2012 speaker trait challenge," in *Proceedings of Interspeech*, 2012.

[86] K. Polzehl, T. ans Schoenenberg, S. Moller, F. Metze, G. Mohammadi, and A. Vinciarelli, "On speaker-independent personality perception and prediction from speech," in *Proceedings of Interspeech*, 2012.

[87] B. Schuller. (2012) The Interspeech 2012 speaker trait challenge. [Online]. Available: http://emotion-research.net/sigs/speech-sig/is12-speaker-trait-challenge

[88] J. Pohjalainen, S. Kadioglu, and O. Rasanen, "Feature selection for speaker traits," in *Proceedings of Interspeech*, 2012.

[89] C. Chastagnol and L. Devillers, "Personality traits detection using a parallelized modified SFFS algorithm," in *Proceedings of Interspeech*, 2012.

[90] A. V. Ivanov and X. Chen, "Modulation spectrum analysis for speaker personality trait recogniton," in *Proceedings of Interspeech*, 2012.

[91] C. Montacie and M. Caraty, "Pitch and intonation contribution to speakers' traits classification," in *Proceedings of Interspeech*, 2012.

[92] M. H. Sanchez, A. Lawson, D. Vergyri, and H. Bratt, "Multi-system fusion of extended context prosodic and cepstral features for paralinguistic speaker trait classification," in *Proceedings of Interspeech*, 2012.

[93] Y. Attabi and P. Dumouchel, "Anchor models and WCCN normalization for speaker trait classification," in *Proceedings of Interspeech*, 2012.

[94] K. Audhkhasi, A. Metallinou, M. Li, and S. Narayanan, "Speaker personality classification using system based on acoustic-lexical cues and an optimal tree-structured bayesian network," in *Proceedings of Interspeech*, 2012.

[95] H. Buisman and E. Postma, "The log-gabor method: Speech classification using spectrogram image analysis," in *Proceedings of Interspeech*, 2012.

[96] J. Wagner, F. Lingenfelser, and E. Andre, "A frame pruning approach for paralinguistic recognition tasks," in *Proceedings of Interspeech*, 2012.

[97] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social Signal Processing: Survey of an emerging domain," *Image and Vision Computing Journal*, vol. 27, no. 12, pp. 1743–1759, 2009.

[98] J. Biel and D. Gatica-Perez, "The Youtube lens: Crowdsourced personality impression and audiovisual of vlogs," *IEEE Transactions on Multimedia*, vol. 15, no. 1, pp. 41–55, 2012.

[99] J. Staiano, B. Lepri, R. Subramanian, N. Sebe, and F. Pianesi, "Automatic modeling of personality states in small group interactions," in *Proceedings of the ACM International Conference on Multimedia*, 2011, pp. 989–992.

[100] R. Srivastava, J. Feng, S. Roy, S. Yan, and T. Sim, "Don't ask me what I'm like, just watch and listen," in *Proceedings of the ACM International Conference on Multimedia*, 2012, pp. 329–338.

[101] M. Rojas, D. Masip, A. Todorov, and J. Vitria, "Automatic prediction of facial trait judgments: appearance vs. structural models," *PLoS ONE*, vol. 6, no. 8, pp. 1–12, 2011.

[102] S. Fitzgerald, D. Evans, and R. Green, "Is your profile picture worth 1000 words? photo characteristics associated with personality impression agreement," in *Proceedings of AAAI International Conference on Weblogs and Social Media*, 2009.

[103] D. Evans, S. D. Gosling, and A. Carroll, "What elements of an online social networking profile predict target-rater agreement in personality impressions," in *Proceedings of the International Conference on Weblogs and Social Media*, 2008, pp. 45–50.

[104] C. Nass and S. Brave, *Wired for speech: How voice activates and advances the Human-Computer relationship*. The MIT Press, 2005.

[105] M. Schmitz, A. Kruger, and S. Schmidt, "Modeling personality in voice of talking products through prosodic parameters," *In Proceedings of the International Conference on The Internet of Things*, pp. 313–316, 2007.

[106] J. Trouvain, S. Schmidt, M. Schroder, M. Schmitz, and W. J. Barry, "Modeling personality features by changing prosody in synthetic speech," *In Proceedings of Speech Prosody*, 2006.

[107] J. Chin, "Personality trait attributions to voice mail user interfaces," in *Conference Companion on Human Factors in Computing Systems: Common Ground*, 1996, pp. 248–249.

[108] E. De Sevin, S. Hyniewska, and C. Pelachaud, "Influence of personality traits on backchannel selection," in *Proceedings of Intelligent Virtual Agents*, 2010, pp. 187–193.

[109] M. McRorie, I. Sneddon, G. McKeown, E. Bevacqua, E. de Sevin, and C. Pelachaud, "Evaluation of four designed virtual agent personalities," *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 311–322, 2011.

[110] M. Neff, Y. Wang, R. Abbott, and M. Walker, "Evaluating the Effect of Gesture and Language on Personality Perception in Conversational Agents," in *Proceedings of the International Conference on Intelligent Virtual Agents*, Philadelphia, 2010, pp. 222–235.

[111] E. Krahmer, S. Van Buuren, and W. Wesselink, "Audio-visual personality cues for embodied agents: An experimental evaluation," *In Proceedings of the Workshop on embodied conversational characters as Individuals*, 2003.

[112] A. Egges, S. Kshirsagar, and N. Magnenat-Thalmann, "Generic personality and emotion simulation for conversational agents," *Computer Animation and Virtual Worlds*, vol. 15, no. 1, pp. 1–13, 2004.

[113] D. Arellano, J. Varona, F. Perales, N. Bee, K. Janowski, and E. Andre, "Influence of head orientation in perception of personality traits in virtual agents," in *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems*, 2011, pp. 1093–1094.

[114] A. Tapus, C. Ţăpuş, and M. Mataric, "User—robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy," *Intelligent Service Robotics*, vol. 1, no. 2, pp. 169–183, 2008.

[115] A. Tapus and M. Mataric, "Socially assistive robots: The link between personality, empathy, physiological signals, and task performance," in *Proceedings of AAAI Spring Symposium*, 2008.

[116] S. Jung, H. Lim, S. Kwak, and F. Biocca, "Personality and facial expressions in human-robot interaction," *Proceedings of the Annual ACM/IEEE International Conference on Human-Robot Interaction*, pp. 161–162, 2012.

[117] H. Miwa, A. Takanishi, and H. Takanobu, "Experimental study on robot personality for humanoid head robot," *Proceedingsof the IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 2, pp. 1183–1188, 2001.

[118] S. Woods, K. Dautenhahn, C. Kaouri, R. Boekhorst, and K. Koay, "Is this robot like me? Links between human and robot personality traits," in *IEEE-RAS International Conference on Humanoid Robots*, 2005, pp. 375–380.

[119] M. Walters, K. Dautenhahn, R. Te Boekhorst, K. Koay, C. Kaouri, S. Woods, C. Nehaniv, D. Lee, and I. Werry, "The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment," in *IEEE International Workshop on Robot and Human Interactive Communication*, 2005, pp. 347–352.

[120] K. M. Lee, W. Peng, S.-A. Jin, and C. Yan, "Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human-robot interaction," *Journal of Communication*, vol. 56, no. 4, pp. 754–772, 2006.

[121] A. Weiss, B. van Dijk, and V. Evers, "Knowing me knowing you : Exploring effects of culture and context on perception of robot personality," in *Proceedings of the International Conference on Intercultural Collaboration*, 2012, pp. 133–136.

[122] J. Howe, "The rise of crowdsourcing," *Wired Magazine*, vol. 14, no. 6, 2006.

[123] D. Schneider, "Implicit personality theory: A review." *Psychological Bulletin*, vol. 79, no. 5, p. 294, 1973.

[124] G. Mohammadi and A. Vinciarelli, "Humans as feature extractors: Combining prosody and personality perception for improved speaking style recognitio," in *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, 2011, pp. 363–366.

[125] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 451–515, 2000.

[126] J. Machajdik and A. Hanbury, "Affective image classification using features inspired by psychology and art theory," in *Proceedings of ACM International Conference on Multimedia*, 2010, pp. 83–92.

[127] R. Datta, D. Joshi, J. Li, and J. Wang, "Studying aesthetics in photographic images using a computational approach," in *Proceedings of the European Conference on Computer Vision*, 2006, pp. 288–301.

[128] M. Tkalcic, T. Tasic, and J. Kosir, "Emotive and personality parameters in multimedia recommender systems," in *Proceedings of the IEEE International Conference on Affective Computing and Intelligent Interaction*, 2009.

[129] M. Tkalcic, J. Tasic, and A. Košir, "The LDOS-PerAff-1 corpus of face video clips with affective and personality metadata," in *Proceedings of the Workshop on Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*, 2010.

[130] M. Aylett, "Cereproc," 2013. [Online]. Available: http://www.cereproc.com

[131] World Economic Forum, "Personal data: the emergence of a new asset class," World Economic Forum, Tech. Rep., 2011.

[132] C. Saint-Georges, A. Mahdhaoui, M. Chetouani, R. Cassel, M.-C. Laznik, F. Apicella, P. Muratori, S. Maestro, F. Muratori, and D. Cohen, "Do parents recognize autistic deviant behavior long before diagnosis? taking into account interaction using computational methods," *PLoS ONE*, vol. 6, no. 7, p. e22393, 2011.

[133] F. Warren, "Treatment of personality desorders," in *The Cambridge handbook of personality psychology*, P. Corr and G. Matthews, Eds. Cambridge University Press, 2009, pp. 799–819.