PhD Dissertation

Doctorate School in Cognitive and Brain Sciences

CIMeC - University of Trento

Multimodal Personality Recognition from Audiovisual Data

Ligia Maria Bătrînca

Advisor: Prof. Dr. Fabio Pianesi

Cycle XXVI
December 2013

Acknowledgements

Who I am today is the sum of all experiences and people I have met throughout my life.

I have reached the point in my life, which I have been thinking about since I started my PhD. I would like to express my gratitude...

To Fabio Pianesi,
For his patience, dedication, teaching me and inspiring me to give my best, throughout these 3 years.

To my friends,
For encouraging me and making me smile. Ailbhe receives an extra big thank you for proof reading this work and telling me to keep calm.

To Louis-Philippe Morency and friends in the MultiComp group, at Institue for Creative Technologies in California, for showing me other exciting research directions. They did so without probably knowing, and I am most grateful to them.

Most importantly, to my family,
Especially my beautiful and loving parents, Aurica and Ion. Everything I am, I owe it to them.

Abstract

Automatic behavior analysis lies at the intersection of different social and technical research domains. The interdisciplinarity of the field, provides researchers with the means to study the manifestations of human constructs, such as personality. A branch of human behavior analyis, the study of personality provides insight into the cognitive and psychological construction of the human being. Research in personality psychology, advances in computing power and the development of algorithms, have made it possible to analyze existing data in order to understand how people express their own personality, perceive others', and what are the variables that influence its manifestation.

We are pursuing this line of research because insights into the personality of the user can have an impact on how we interact with technology. Incorporating research on personality recogniton, both from a cognitive as well as an engineering perspective, into computers could facilitate the interactions between humans and machines. Previous attempts on personality recognition have focused on a variety of different corporas (ranging from text to audiovisual data), different scenarios (interviews, meetings), different channels of communication (audio, video, text) and different subsets of personality traits (out of the five ones present in the Big Five Model: Extraversion, Agreeableness, Conscientiousness, Emotional Stability and Creativity). Our work builds on previous research, by considering simple acoustic and visual non-verbal features extracted from multimodal data, but doesn't fail to bring novelties: we consider previously uninvestigated scenarios, and at the same time, all of the five personality traits and not just a subset.

In the first part, we look closely at a self-presentation scenario and what it can reveal in terms of ability to capture emerging personality traits. Second, we look at the Human-Computer Interaction scenario. In this scenario we introduce another novelty of our work: the display of different "collaboration levels", ranging from fully-collaborative to fully non-collaborative during the interaction with the subject. Finally we look at the contribution of the third scenario, Human-Human Interaction, on the emergence of personality traits. Investigating this scenario creates a much stronger basis for future human-agents interactions.
Our goal is to study the degree emergence of personality traits in these three different scenarios: self-presentation, human-computer interaction and human-human interaction. The results demonstrate the relevance of each of these three scenarios, when it comes to the degree of emergence of certain traits.

Keywords Personality, behavior analysis, self-presentation, human-machine interaction, human-human interaction, nonverbal cues, automatic personality inference, map task.

Contents

List of Figures

List of Tables

1. Introduction

1.1 Motivation

The increasing pace of systems and applications development shifts the users' focus on two important notions: adaptivity and usability. Both are important to address in order to achieve a more human-like interaction with machines. Being able to recognize the user's personality and react to it can have many advantages for different applications, such as multi-modal interaction based systems, computer supported collaborative work, intelligent tutoring systems (user and system behavior modelling) or in many fields, such as ubiquitous computing and forensics, as the analysis of textual excerpts might offer valuable insight about the personality of the note's author [75].

Studies have shown that our attitute towards machines is also influenced by our personality [105]. In [95], the authors point out that users tend to treat computers as real people, and the evaluation of conversational agents depends on their own personality [24]. Therefore, the next generation of systems that display a more human-like behavior and are endowed with the ability to adapt to the user's personality, represent a requirement and an emerging research challenge.

Taking intelligent tutoring systems as an example, it has been shown that the learning process and the motivation behind it, are correlated with the personality dimensions of the Big Five model. In [63] the authors investigate the relation between the Big Five traits, academic motivation, and academic achievement at the same time, and conclude that the conscientiousness traits is central to all three types of academic motivation (cooperative, hypercompetitive, and personal). Personality factors also account for children's reading motivation [77]. Another effect of developing and providing students with personality-adaptive tutoring systems, can be that the learning experience becomes a positive one, thus encouraging future interaction with the system.

An analysis of the effects of personality on leadership, done by Hogan et al., [54], proved that there is a strong relationship between the ability to exercise leadership and personality. Also when it comes to the sales domain and job performance, personality traits play a major role [45]. Automatic personality recognition employed in these domains can, for instance, improve the outcome of the identification process for selecting the most suitable candidate for the job.

In their book, the authors Byron Reeves and Clifford Nass [95], also mention another field that can benefit from recognizing the personality of the user, in a fast and automatic manner: the advertising industry. Given the nature of the additional information about their customers, companies can tailor the commercial content according to each customer.

Virtual environment applications can also benefit from recognizing the personality of their users. This emerging technology has been used in various training scenarios and psychotherapy (fear of flying [100], [83], public speaking fear [13]). Providing feedback, customized to each subject's personality, can greatly improve the outcome of the therapy session.

As presented above, modeling and recognizing personality has the potential to improve and enhance our interaction experience with computers or applications. Computational recognition of personality is still in its infancy, but the results are promising, thus encouraging further research.

This is how we arrive at the challenge of extending the current state-of-the-art literature on personality recognition.

The work presented in this thesis, and that focuses on recognizing personality from audiovisual data, is motivated by two important aspects. First, we envision the advantages of a system capable of assessing the user's personality based on the fusion of acoustic and visual signals. Second, the three scenarios considered in this work, self-presentation, goal-oriented human-computer interaction and human-human interaction, are to the best of our knowledge addressed for the first time.

The first scenario is relevant if we consider the way personality affects how people convey their image and how others perceive first-impressions. This is an important topic in many social computing domains. The other two studies, dealing with the emergence of the five personality traits from audiovisual data, in interactive human-computer and human-human interaction scenarios, are an attempt to complement the existing picture of personality recognition. Our findings can be extented to the adjacent research field of personality generation (systems or virtual characters).

As with any work of inquiry, the emerging scientific questions that this work pursues distinguish it from other studies and makes it an important milestone for future work.

1.2 Personality Definitions

The psychological concept of personality has multiple definitions and conveyed different meanings. A few personality definitions, in chronological order, are given below:

> *"the dynamic organization within the individual of those psychophysical systems that determine his characteristic behavior and thought."* [2]

“The collective perceptions, emotions, cognitions, motivations, and actions of the individual that interact with various environmental situations.” [32]

”An individual’s pattern of psychological processes arising from motives, feelings, thoughts, and other major areas of psychological function. Personality is expressed through its influences on the body, in conscious mental life, and through the individual’s social behavior.”[76]

Personality, with its underlying causes and external manifestations, has been on the mind of scholars for longer than one might think. In the 17th century, Franz Joseph Gall, the founder of phrenology, claimed that personality is strongly linked to one’s head shape. This is not the only example when it comes to linking physiological constructs to the underlying mechanism of someone’s behavior. In ancient Greece, Hippocrate proposed, what would become one of the earliest attempts to study personality: The Four Temperaments. It was believed that the balance of four humors (Latin: humor, ”body fluid”) influenced human personality. The four humors are: blood, phlegm, yellow bile and black bile. According to the theory, blood was connected to a temperament known as sanguine (cheerful, even-tempered, confident, optimistic), phlegmatic (shy, unemotional and relaxed), choleric (hot-tempered, dominant, leader material, ambitious) and melancholic (kind, creative, sensitive). This categorization also represents an early attempt of connecting personality to physiological constructs.

Future theories of personality have stemmed from these early attempts to provide a scientific explanation for a theoretical notion.

1.3 The Big Five

The systematic study of psychology has produced a number of particular views with respect to how people differ in their behavior, making an attempt to determine the underlying causes and processes of these differences. Studying how these different processes integrate with other deterministics or causal elements, to give each person a distinctive identity, also contributed to the theories of personality that are known today.
Below we will introduce in a few words some of the major theories of personlity, often mentioned in psychology and personality literature.

The trait theory
Psychologist who support the trait theory, acknowledge that personality is the manifestation of specific traits. A trait is a stable characteristics, meaning it is constant throught someone's life. Representatives of the trait theory were Gordon Allport and Hans Eysenck.

The biological theory
The followers of this theory emphasize the influence of genetic factors, to a certain degree, on the manifestation of personality. In [20], the authors provide an extensive review of literature regarding the biological theory of personality

The behavioral theory
The supporters of this theory emphasize the relevance and importance of the interaction between an individual and their surrounding environment, when determining its personality R. The theory is also known as behaviorism. Among many behavioral theorists, we can include Benjamin Skinner and John Watson.

The psychodynamics theory
This theory is based on the work of Sigmund Freud. It encompasses the idea that personality is the "interaction" between different layers: id, ego and superego.

The humanist theory
This theory puts the emphasis on free-will. Self-actualization is seen as a way to constantly improve and fulfill oneself. The followers of this theory reject the biological determinism and the behaviorist view of conditioning. To pursue our research goal we count on the ability to identify and capture direct or indirect manifestations of personality. The approach is very close to the definition of trait theory, and therefore we adopt this theory throughout our work. With respect to the trait approach to personality theory, the labels "Big-Five" and "Five-Factor Model" are often interchangeably by researchers. The Big-Five and Five Factor Model provide a common taxonomy of personality traits [59].

The five-factor structure, as it is known today, began from the work of psychologist Raymond Cattell. He developed the 16 Personality factors questionnaire [51]. Since then, personality scholars have dedicated decades of additional research to refining and validating Cattell's model. Psychologist Donald Fiske further improved Cattell's model, shaping what would later be known as the Big Five and provided his model as input for Ernest Tupes and Raymond Christal's further factor analyses in 1961 [60]. Finally, Costa and McCrae published the revised version of the NEO Personality Inventory, which allows for the measurement of each Big Five dimensions [29]. Table 1.1 introduces the five components and the high-level descriptors for each trait.

Although researchers do not agree on the exact label of the factors, there is a consensus that the behavior represents the interaction between

Table 1.1: Costa and McCrae Big Five Dimensions

Factors	Description
Extraversion	The core feature of extraversion is the propensity to seek and hold social attention [10]. High scorers tend to be outgoing assertive, while low scorers are perceived as reserved and shy.
Neuroticism	Children and adults who are high on the neuroticism spectrum, tend to experience the surrounding environment in a distressed mode, exhibit lack of confidence and be insecure, while low scorers are more emotionally stable [23].
Conscientiousness	Conscientious people are characterized by self-discipline, organized manner and are more willing to follow authority [53]. Those low on this trait are perceived as inefficient and careless.
Agreeableness	Agreeable people take interest in pleasuring their peers, they are friendly and cooperative. Opposite to this are antagonistic and faultfinding people.
Openess to Experience	This higher-degree trait is characterized by imaginative, insightful and creative. The opposite is shallow, unimaginative.

a person's personality and different variables, such as gender, culture or social context. An extended review of the literature of gender differences in personality and across cultures is provided by [41] and [30].

We chose to further work with the Big Five, because its cross-culturality [102] and its longitudinal consistency [82] has been pointed out. The trait domains of the Five Factor Model offer a complete and ranged image for describing human behavior and is broad enough to be used in computational science as a measure of human behavior. The aim of our work is not to capture finer distinctions of behavior and so a broad model of personality is appropriate for our work. Another reason to use this model was that the Big Five has been validated in the Italian language [92] and this is an important aspect, since our data is in Italian language.

1.4 Machine Learning Algorithms

This section provides an overview of the machine learning algorithms used in the classification experiments and its purpose is to give the readers with different machine learning knowledge a starting point in understanding the work described in this thesis. By no means is this intended as a deep analytical mathematical illustration. For an indepth analysis of Support Vector Machines, we direct the reader to [28]. While Witten and colleagues provide a simple and elegant explanation of Naïve Bayes classifier in their book "Data Mining" [108], a good resource is also the book by Cristopher Bishop, "'Pattern Recognition and Machine Learning.' [16].

1.4.1 Support Vector Machines

The support vector machine (SVM) [28] is an established model of data in machine learning. In this work, we learn SVMs for binary classification, that is to assign one of the two labels, Low (L) or High (H), to a given instance (example), for any of the five personality traits, components of the Big Five Model. We used linear SVMs as well as nonlinear SVM with a Radial Basis Kernel.

Linear SVM

A support vector machine is a classifier that, given labeled training instances (supervised learning), outputs a hyperplane that separates new given instances.

For instance, given a set of n input vectors $\mathbf{x}_i$ and with possible outputs $y_i \in \{-1, +1\}$, the goal is to find the optimal hyperplane that correctly separates the instances belonging to the two classes. Where the dimension of the input vectors is 2, the problem is reduced to finding the optimal line. If the dimension is 3, then it is a plane and if the dimension is greater than 3, then we must find the optimal hyperplane.
The equation for the separating hyperplane is:

$$\mathbf{x} \cdot \mathbf{w} + b = 0 \tag{1.1}$$

where $\mathbf{w}$ is the normal vector to the hyperplane.

Let's take a step back and simplify things a bit. In Figure 1.1 we illustrate an example of linearly separable data. Several candidate lines separate the points that belong to the two classes. The only question is: how do we determine the line that will better separate future, unseen instances into the two classes?

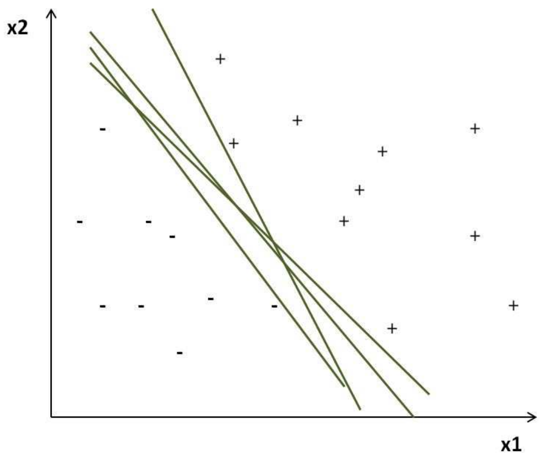


Figure 1.1: 2-Dimension Separable Data

This question leads us to the steps necessary for implementing an SVM. The implementation of an SVM for linearly separable data is reduced to determining the vector $\mathbf{w}$ and parameter b from Equation 1.1 so that the data can be separated according to the equations:

$$\mathbf{x}_i \cdot \mathbf{w} + b \geq +1; \quad for \quad y_i = +1 \tag{1.2}$$

$$\mathbf{x}_i \cdot \mathbf{w} + b \leq -1; \quad for \quad y_i = -1 \tag{1.3}$$

To start with, the optimal hyperplane is determined in terms of its ability to generalize. To do so, it must not be too close to the training points in either class. It follows that the distance to these points must be maximized. The double of this distance is called the *margin*. We can say that the optimal hyperplan maximizes the margin of the training data. Figure 1.2 presents graphically the margin and the training points that determine the margin separators and are called **support vectors**.

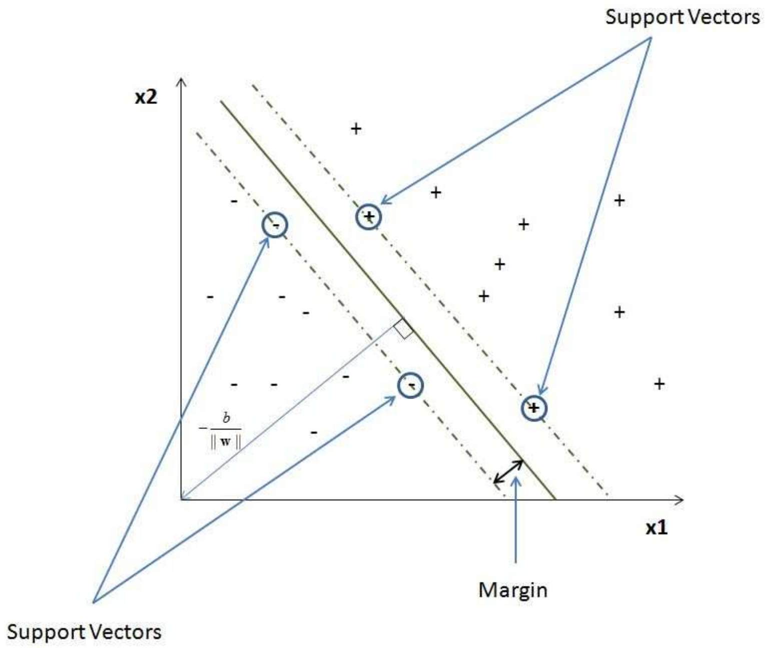


Figure 1.2: A hard margin hyperplane for linearly separable data.

Through a series of mathematical steps, as previously mentioned, vector $\mathbf{w}$ and parameter b are computed in order to maximize the margin of the support vector. In the context of linearly separable data, this translates into minimizing $||\mathbf{w}||^2$ subject to Equations 1.2 and 1.3. A complete analysis is provided in [28].

When it is not possible to determine a plane that completely separates the points from the two classes, it means we are dealing with non-separable data. In this case, we can extend the margin to allow non-separating planes, or for data to be misclassified. To serve this purpose is the *slack variable* ξ_i. Although some instances will be misclassified, they will be penalized. Introducing the slack variable, ξ_i, equations 1.2 and 1.3 become:

$$\mathbf{x}_i \cdot \mathbf{w} + b \geq +1 - \xi_i; \quad for \quad y_i = +1 \tag{1.4}$$

$$\mathbf{x}_i \cdot \mathbf{w} + b \leq -1 + \xi_i; \quad for \quad y_i = -1 \tag{1.5}$$

$$\xi_i \geq 0 \forall i \tag{1.6}$$

In this case, we want to minimize $||\mathbf{w}||^2 + C \sum_{i=1}^{n} \xi_i$, subject to Equations 1.4 and 1.5. The scope of parameter C is to tune the trade-off between the size of the margin and the penalty of the slack variable. A large value for C corresponds to assigning a higher penalty to the classification errors. Figure 1.4 presents all an example of non-separable data, that illustrates the concept of soft margin.

Nonlinear SVM with RBF

When data is not linearly separable (cannot be classified by a linear classifier) another approach must be adopted. A solution to this problem is to project the input space $\mathbf{x}$ onto a higher dimensional feature space ("operation" called "the kernel trick"), and then use a linear classifier in the higher dimensional space. The resultant classifier should still generalize well. Let's consider the Linear Kernel:

$$K(\mathbf{x}_i, \mathbf{x}_i) = \mathbf{x}_i^T \mathbf{x}_j \tag{1.7}$$

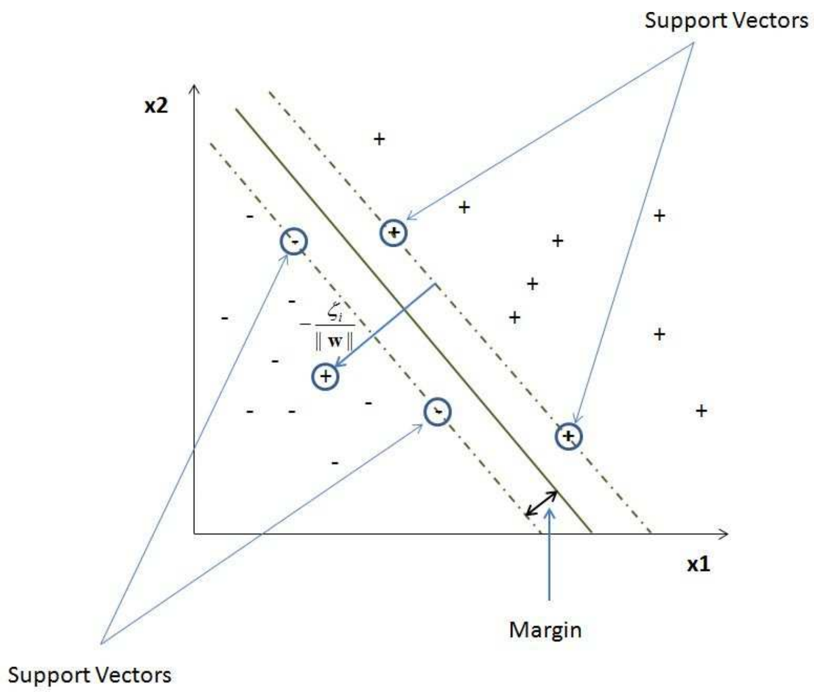


Figure 1.3: A soft margin hyperplane for non-separable data.

Mapping every data point into a higher dimensional space given the transformation:

$$\Phi\colon \mathbf{x} \mapsto \phi(\mathbf{x}) \tag{1.8}$$

the dot product from Equation 1.7 becomes:

$$\phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) \tag{1.9}$$

Kernel functions are merely functions that calculate the dot product of two vectors. The two commonly used families of kernels are polynomial kernels (linear kernels are a subclass of polynomial kernels) and radial basis functions. Radial Basis Function kernels have the form:

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-||(\mathbf{x}-\mathbf{x})^2||/(2\sigma^2)} \tag{1.10}$$

Simply put, a radial basis function allows features to draw circles (spheres). A thorough analysis of Support Vector Machine and kernel functions can be found in [22]. An example that illustratess 2D nonlinear data, mapped using a Radial Basis Kernel [1] into a 3D space.

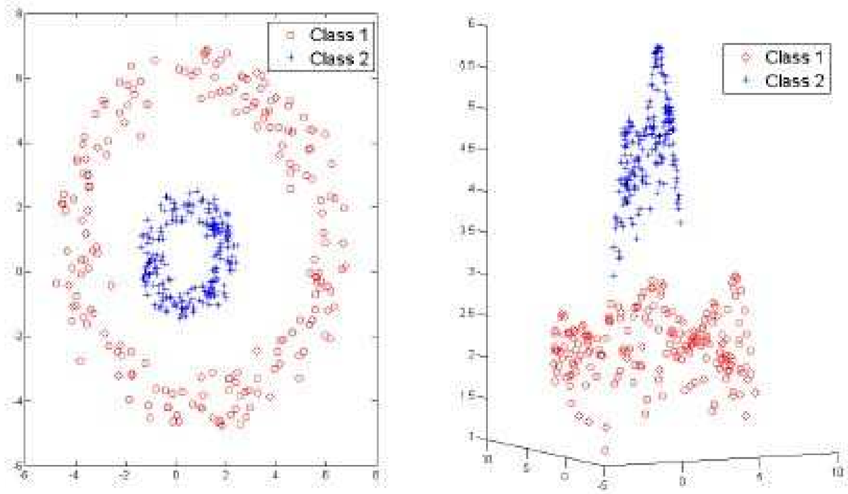


Figure 1.4: Two-dimensional, nonlinear, data re-mapped using Radial Basis Kernel into a three-dimensional space, where it can be classified easier.

1.4.2 Naïve Bayes

Naïve Bayes is a simple probabilistic classifier that applies the Bayes theorem:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \tag{1.11}$$

where $P(B|A)$ is the posterior probability of event B after evidence has been observed; $P(A|B)$ indicates the likelihood of A given B; $P(B)$ is the prior probability of event B before the evidence is observed and $P(A)$ is the evidence we have with respect to event A.

The classifier assumes that the presence or absence of a particular feature of a given class (e.g. a personality state L/H) is unrelated to the presence or absence of any other feature. In this case we say feature x is independent. The main advantage of using Naïve Bayes is that it only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification.

Let $X = (x_1, ..., x_n)$ be the feature vector of an instance I and C the set

of *classes* (or categories), $C = (c_1, ..., c_m)$ that we want to classify instance I in. In our case $C \in \{L, H\}$. We want to determine the probability of I belonging to one of the two categories in C, i.e. $P(C_m|I)$.
Since we know the feature vector X of instance I, we use Bayes' rule (Equation 1.11 to write down the probability:

$$P(c_i|x_1, ..., x_n) = \frac{P(x_1, ..., x_n|C_m)P(c_i)}{P(x_1, ..., x_n)} \tag{1.12}$$

Using the independence assumption:

$$P(x_i|c_i, x_1, ..., x_{i-1}, x_{i+1}, ..., x_n) = P(x_i|c_i) \tag{1.13}$$

we obtain:

$$P(C_i|x_1, ..., x_n) = \frac{P(c_i) \prod_{i=1}^{n} P(x_i|c_i)}{P(x_1, ..., x_n)} \tag{1.14}$$

We can write:

$$P(c_i|x_1, ..., x_n) \simeq P(c_i) \prod_{i=1}^{n} P(x_i|c_i) \tag{1.15}$$

since $P(x_1, ..., x_n)$ is constant. Therefore

$$\hat{c_i} = \arg\max_{c_i} P(c_i) \prod_{i=1}^{n} P(x_i|c_i) \tag{1.16}$$

and we can use a Maximum A Prioriation to estimate $P(c)$ and $P(x_i|c_i)$.

1.5 Thesis Outline

The remainder of this thesis is organized as follows. Chapter 2 surveys existing literature on automatic personality recognition, using different channels of communication, discussing the possible channels of signal capturing. Chapter 3 introduces in a detailed fashion the data we use for our three studies. Chapter 4 presents the implemented methodology to the problem of automatically recognizing personality in a self-introduction setting, using nonverbal cues extracted from audiovisual data. We systematically investigate a top-down approach to determine the best feature set. Chapter 5 illustrates the second setting, Human-Computer Interaction. In order to be consistent and provide a quality results comparisson, the approach taken in this setting, is similar to the previous chapter. In Chapter 6, we examine personality recogniton in the third setting:Human-Human Interaction. Chapter 7 concludes with a summary, an extended discussion about the results presented in previous chapters, the limitations of the work and an outlook on future research directions.

1.6 Thesis Contributions

The content presented in Chapter 4 and 5 has been published in different conference proceedings and workshops. At the time of writing this doctoral thesis, the work presented in Chapter 6 has not been previously published and is currently being prepared for a journal submission.

1. L. Batrinca , N. Mana , B. Lepri , F. Pianesi and N. Sebe "Please, tell me about yourself: Automatic assessment usingshort self-presentations", Proc. Int. Conf. Multimodal Interfaces (ICMI-MLMI), 2011.

2. L. Batrinca, B. Lepri, and F. Pianesi. Multimodal recognition of personality during short self-presentations. In Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding, pages 27–28. ACM, 2011.

3. L. Batrinca, B. Lepri, and F. Pianesi. "Multimodal recognition of personality traits in human-computer collaborative tasks", Proc. Int. Conf. Multimodal Interfaces (ICMI-MLMI), 2012.

2. Literature Review

This research intersects the research areas of computer science and personality psychology. This chapter surveys relevant work on automatic personality recognition and discusses the relevant term of *thin slices* and its importance in human behavior analysis and modelling.

2.1 Personality in Computational Science

A vast body of literature can be found on research done jointly by psychologists and computer scientists on facets of human behavior or the results of its expression in individuals, such as emotion [90], mood [79], mood expressed in song lyrics [66], sentiment analysis [73], dominance [58] or deception [49]. First, studies focused on the influence of personality on the daily, common activities of people or their interactions with computers and their peers. Later, computer scientists teamed with personality psychologists, in an effort to recognize the user's personality from a variety of activities: interaction with work colleagues, online profiles (professional, dating, social networks), speaking or writing style.
The interest in personality has been maintained by the fact, proven on many occasions, that the manifestation of personality can be recognized from a variety of communication channels, such as visual and acoustic. In Figure 2.1 we have an illustrative example of how nonverbal communication can influence the messages that are transmitted and conveyed.

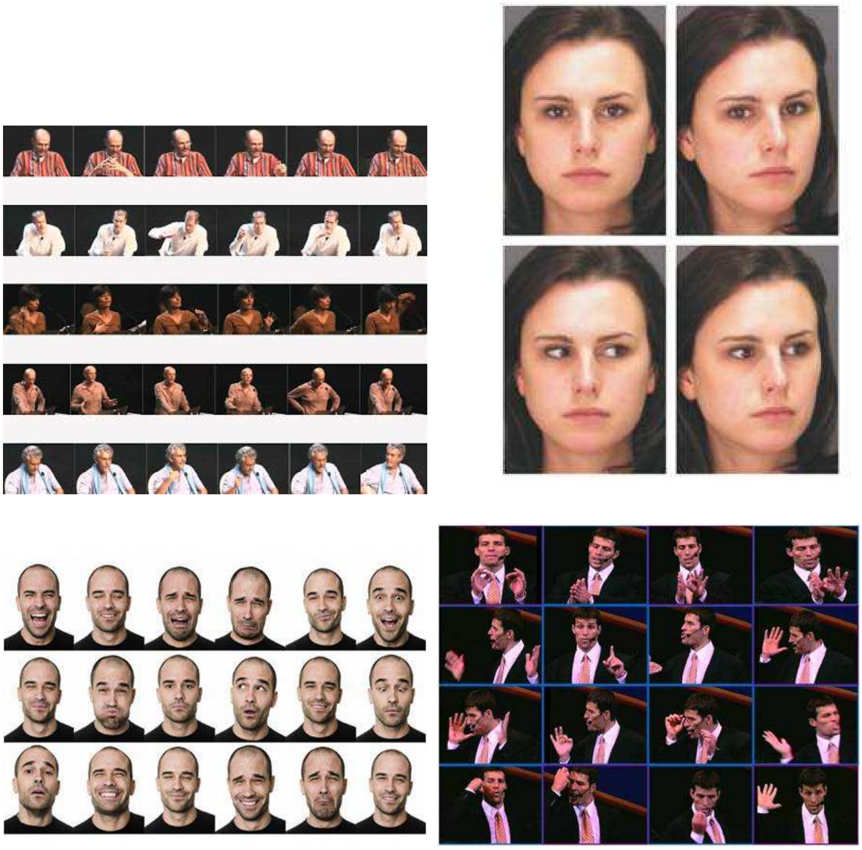


Figure 2.1: From top left, clockwise: examples of how postures, head orientation and eyegaze, hand gestures and facial expressions are used in communication.

Work on automatic recognition of personality is relatively recent. Pioneering work addressing this issue was carried out by Argamon and colleagues [8] in 2005. They focused on just two (Extraversion and Emotional Stability) of the Big Five model components, measured by means of self-reports. Support vector machines (SVMs), trained on word categories based on systemic functional grammar (SMG) and relative frequency of function words, were used to recognize these two traits. Similarly, Oberlander and Nowson [88] worked on automatic classification of author personality to detect four (Agreeableness, Conscientiousness, Extraversion and

Neuroticism) of the Big Five traits on a corpus of personal weblogs. They explored both Naive Bayes classifiers and SVMs, trained on different set of n-gram features. Mairesse and colleagues [75] also worked on recognition of all Big Five personality traits in both conversation and text. According to psycholinguistic and psychosocial literature, they systematically investigated the usefulness of different sets of acoustic and textual features, extracted from self-reports and observer ratings of personality data. They experimented with classification, regression and ranking models, with the latter performing the best. The results showed that: (a) prosodic features play a major role in predicting personality, (b) Extraversion is the easiest personality trait to model from spoken language; and (c) the automatic recognition was closer to observed personality than to that emerging from the self-reports.

part from the written text, researchers also looked for the expression and the recognition of personality in other environments, interactive and non-interactive. For instance, social settings are widely investigated, because of they provide a unique opportunity to observe people, in an almost unobtrusive manner, in their own environment. This is important because it can be done, without subjecting individuals to major constraints.

Olguin et al. [89] collected various behavioral measures, extracted from daily activities of 67 professional nurses in a hospital. The data were collected using sociometric badges, a wearable device integrating a number of sensors (accelerometer, infra-red sensor and microphone) measuring aspects such as physical and speech activity, level of proximity to relevant objects - people, beds - number of face-to-face interactions with others, and social networks parameters. The correlation analysis conducted by the authors showed that personality traits can be extracted from features derived from low-level sensor data. For example, people who scored high for neuroticism, had higher daily percentages of face-to-face (f2f) time. At the same time, less average time the subject spents in close proximity to a bed or phone, the more extrovert they are More recently, Pianesi et al. [93] and Lepri et al. [69] showed the feasibility of automatically recognizing the Extraversion and Locus of Control personality traits in social interactions, using simple non-verbal features. Their approach was based on the assumptions that personality is exhibited in the course of social interaction and thin slices of social behavior are enough to classify personality traits. The first assumption was realized by exploiting classes of acoustic features encoding specific aspects of social interaction (Activity, Emphasis, Mimicry, and Influence) and three visual features (head, body, and hands fidgeting). For the second, they demonstrated personality assessment based on inferences from 1-minute-long behavioral sequences. Proximity to other people and the expression of personality traits has been investigated in [111]. The authors employed proxemic features (e.g., number of intimate, personal, and social relationships; minimum distance between two subjects, etc.), extracted from video tracking and head pose estimation, to investigate the automatic recognition of Extraversion and Neuroticism.

The connection between how users portray themselves in the virtual world, on social networks and their personality has caught the attention of many researchers. A study by [103] examined the relationship between the Big Five and the use of Facebook to fulfill self-presentational needs. Self-presentational behaviors and motivations were best predicted by low conscientiousness and high neuroticism. Results suggest that conscientious individuals are cautious in their online self-presentation. Neuroticism, agreeableness, and extraversion were positively associated with the tendency to express ones actual self. Another study by Kosinski and colleagues [64] looked at the predictive power of Facebook Likes to infer personality traits and other personal attributes such as sexual orientation, ethnicity, religious and political views.

How personality is used in online dating services, has been mentioned in [39]. Although, the main work is on self-presentation strategies among online dating participants, who wish to find a romantic partner, what is interesting is that participants prefer to actually depict aspects of their personality in their profiles and not hide it. They even present content (personal stories, pictures) which confirms their personality traits.

Smartphones provide functions that allow the owner to e-mail, go online (check their stock market portofolio, calendar, news, video conferencing via Skypeect.) or play music and games (entertainment). The growth of smartphone usage provides a whole new data for researchers who are interested in looking at what lies beyond the user's phone usage owner: their personality. Several studies have looked at how people accept, and use smartphone technology. In [65], the authors, investigated which functions are most important for people in all five traits. They found that extraverts prefer texting to calling, which is supported by previous research done by [38], while users who score high on Agreeableness, prefer calling to texting. Ehrenberg [38] conducted a study with 200 university students, examining the role of personality and self-esteem in mobile phone use. Their findings showed that neurotic individuals spend more time writing text messages and are prone to stronger mobile phone addictive tendencies. Recently, de Oliveira et al. [34] showed that variables derived from the users mobile phone call behavior as captured by call detail records and social network analysis of the call graph can be used to automatically infer the users personality traits defined by the Big Five. On the same line, Chittaranjan et al. [26] analyzed the relationship between smartphone usage and self-assessed personality. Their study is based on a large-scale dataset of 17 months of real usage of smartphones by 117 people and personality surveys that are suitable for large mobile or online studies. Application usage, call and SMS logs contained several meaningful relationships to the Big-Five

personality framework. The authors also point out the necessity to aim for gender-specific models when predicting personality.
Personality is affecting the way we experience the environment. The association between positive emotions and the five personality dimensions have been studied and reviewd in [104]. The results suggested that Extraversion, Conscientiousness, Agreeableness and Openess to Experience are associated with joy. Lucas et al., [74] showed that extraverts have higher positive affect than introverts across a multitude of situations. They enjoy social activities as much as extraverts but they are not as aroused by it, as Extraverts.

2.2 Thin Slices

It is common to make assumptions about someone, especially when we first meet that person. It's a characteristic developed and used everyday by everyone. This is connected to our instinct to quickly assess a situation in order to determine whether it represents a threat or not. Among other factors, this is what helped our species to survive. Although previous studies showed empirical observations of the phenomena that humans are able to make correct guesses with respect to other aspects/facets of their human peers, it was Nalini Ambady and Robert Rosenthal [4] who coined the term of "*thin slices*". This notion is merely a reference to the short amount of expressive, nonverbal behavior people need to understand cues regarding someone's personality and intelligence [19], opinion about someone's sexual orientation [3] or to predict end-of-semester student evaluations of teachers [5] or outcomes of negotiations [33].

Two established tests, namely the Profile of Nonverbal Sensitivity (PONS) and the Interpersonal Perception Task (IPT), show the potential of thin slices. The first test, PONS, developed by Rosenthal and his colleagues in 1979, is a test that measures the accuracy of external judges in decoding affective nonverbal cues from face, body, and voice and combinations of these features [98]. The PONS test has been validated cross-cultur. It is a collection of 220 videos, each lasting 2 seconds. The video fragments are extracted from the longer video recording, in which a woman portrays 20 different social situations. The findings were astonishing: after removing the verbal content in the videos, the accuracy of the examiners' judgements were better than chance. The second test, the Interpersonal Perception Task (IPT) studies the process of social perception [31]. The goal of this method is to study five types of social interaction: status, level of romantic relationship, kinship, competition and deception. External reviewers looked at 30 different, mundane scenes, all containing unaltered behavior. Each scene had a varied duration, anywhere between 28 and 124 seconds. Results confirmed previous findings: given the presence of nonverbal behaviors, the accuracy rates exceeded chance.

The informative power of thin slices is very important if we take into consideration the goal we envision for future interactions with computers. We expect an improved interaction, something that levels human-human like interactions. For this to happen, we must insert real-time feedback and reaction as part of the interaction equation. Computing power becomes faster and cheaper and this allows us to process and delve into a huge amount of data to extract information that will help us understand better constructs such as personality. This will bring us one step closer to our envisioned course.

Based on these findings, we decided to explore the advantages of using information from thin slices, for the task of recognizing personality in different, and complementary scenarios.

3. Personality Corpus

For the purpose of our research, to study of emergence of personality traits in different scenarios, we designed a tailored dataset, which consists of a collection of video recordings. Each video is comprised of three different parts recorded consecutively in one take. A pause clearly marks each component's part end and the beginning of the next.

The first part of the video is the self-presentation session, where subjects had to introduce and present themselves in just a few words. The total number of self-presentation sessions represent the "Self-Presentation corpus". The second part represents our modified version of the original Map Task corpus [6]. Simply put, the subject's task is to guide someone using a map, from the start point to the end. The "Map Task corpus" is comprised of all these sessions. The third, and last part, is the conclusion and consists of the subject expressing their opinion about the interaction and the task.
In the following sections we will provide a detailed presentation of the entire data collection. Due to our research goals, we will take into consideration and present only the first two parts: the Self-Presentation and Map Task corpus.

3.1 Self-Presentation Corpus

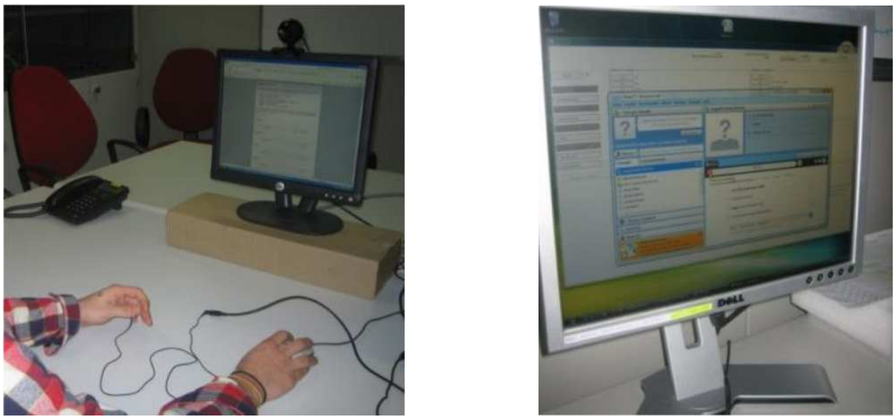
This section will detail the technical setup and recording procedure of the self-presentation data, the task description and statistics about the participants.

3.1.1 Technical Setup and Recording Procedure

The subjects were invited to sit in front of a computer monitor with a webcam on top (see Figure 3.1(a)). At the beginning of the session the participants were asked to sign the informed consent form, to fill in an online personality questionnaire (see section 3.1.4) and they were also provided with the necessary information about the task. After checking that the subjects were wearing the clip-on microphone properly and securing that the webcam was properly placed for frontal and central shots, the experimenter left the participant's room and went to the recording room. In the recording room, the experimenter used a very similar setup (see Figure 3.1(b)), but using only the microphone to communicate with the subject, and not the webcam. From the recording room, the experimenter called the subject via Skype, informing them that they could start the self-introduction session, whenever they were ready.

3.1.2 Task Description

The subjects were asked to introduce themselves in front of a camera. Possible topics (e.g. talking about their job, the book they last read, last holiday, preferred food, preferred sport, etc.) were suggested. However, they were left free to choose any of the suggested topics or any other topics of their choice. In order to obtain a good quality frontal image of the subject to later be used for initialization purposes, the participant was initially asked to look into the camera for a few seconds without moving.



(a) View of subject's experimental setting (b) View of the experimenter work station

Figure 3.1: Experimental Setup

After this brief step, the self-introduction session started. The length of the resulting self-presentations ranged from 30 up to 120 seconds.

3.1.3 Participants

The 89 participants of our study were recruited among employees of a research centre (9 subjects from administration and 14 subjects from various research areas), university students (43 subjects) and other external people (23 subjects). The distribution of the participants was quite balanced in terms of gender (46 male and 43 female) and age (47 young people, i.e. under 25, and 42 adults, i.e. over 25). Due to the participation of many students, the average age was quite low (29 years) but in any case falling within the adult range (over 25). Frontal snapshots of some participants are shown in Figure 3.2.

3.1.4 Personality Questionnaire and Scores

Before the start of the experimental session we asked the participants to fill in the Italian version of the Big Five model personality questionnaire, validated in the Italian language [92]. In this version each trait is investi-

Figure 3.2: Snapshots of participants.

gated through ten items, each of them assessed by means of 1 (completely disagree) to 7 (completely agree) point Likert scale. The procedure used to calculate the Big Five scores is detailed in the Appendix. The results (average and standard deviation values) are reproduced in Table 3.1.

Table 3.1: Averages and standard deviations of Big Five Scores

	Women		Men	
	<25	≥ 25	<25	≥ 25
Agreeableness	49.05	50.68	47.65	51.45
	6.383	5.867	7.985	7.564
Conscienciousness	41.6	46.36	39.45	48.25
	9.643	8.693	9.944	9.124
Emotional Stability	36.57	36.59	40.50	44.65
	7.047	5.518	7.458	6.968
Creativity	45.43	46.55	48.69	47.00
	6.896	7.482	6.279	7.861
Extraversion	47.00	45.50	42.04	42.55
	7.791	8.245	9.075	8.370

In order to investigate the dependence of traits scores on gender and age, the latter was split in two classes: people younger than 25 and people 25 or more years old. The data was examined by means of a series of Analysis of Variance (ANOVA) with the factorial scores for Extraversion, Agreeableness, Conscientiousness, Emotional Stability and Creativity as dependent variables, and Age and Gender as predictors. The results are reported in Table 3.2.

Table 3.2: Dependence of personality traits scores on subjects gender and age. F-statistics values for main and interaction effects from ANOVA analyses. *p<.05; **p<.01; ***p<.001; n.s. - not significant

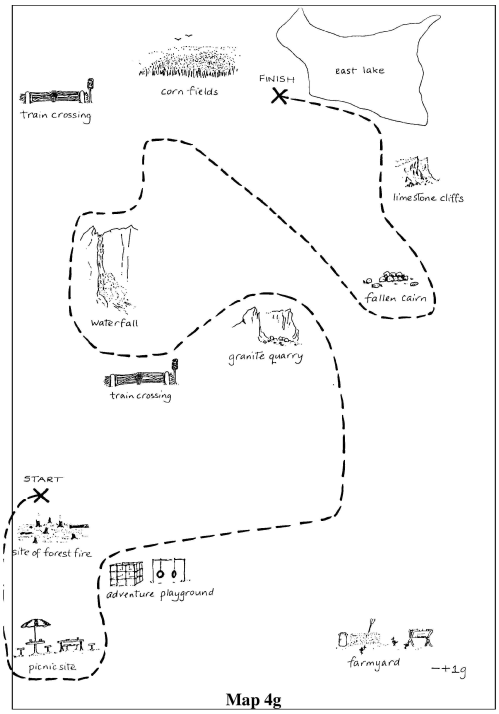
	Gender	Age	Gender*Age
Agreeableness	n.s	n.s	n.s
Conscienciousness	n.s	11.387***	n.s
Emotional Stability	17.054***	n.s	n.s
Creativity	n.s	n.s	n.s
Extraversion	4.859*	n.s	n.s

Evidenced in Table 3.2, no effects were found for Agreeableness and Creativity. Conscientiousness exhibited an age effect (F=11.387, $p<.05$) and Emotional Stability and Extraversion a gender effect (F=17.054 and F=4.859 respectively, $p<.05$). According to our data, people tend to become more conscientious with age (47.26 vs. 40.43) independently of their gender, while on the other hand, men are more emotionally stable than women (42.30 vs. 36.58), whereas women tend to be more extravert (46.23 vs. 42.26).

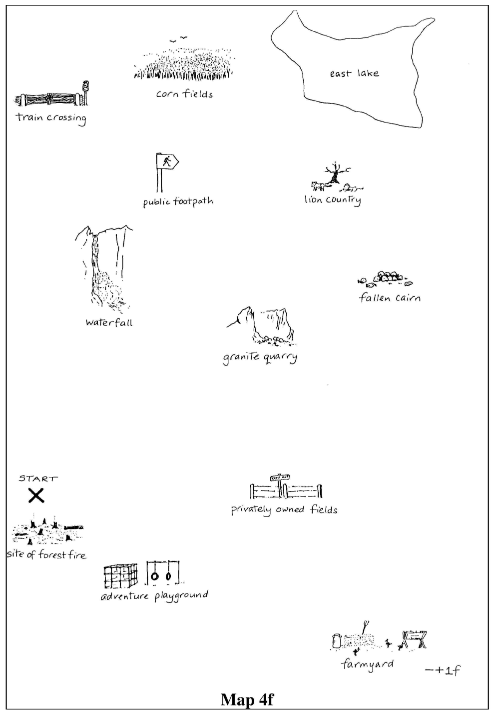
3.2 Map Task Corpus

The original Map Task Corpus [6], introduced by the Human Research Communication Center (HRCR) group at Edinburgh University in 1991, is a cooperative task widely used in psychology, sociology and linguistics studies to collect spontaneous dialogues and interactions, elicited by a task solution.

In the original task two speakers sit opposite one another, separated by a panel. Each of them has a map, not visible to the other. One speaker, designated as “Instruction Giver”, has a route marked on their map (see Figure 3.3(a)); the other speaker, designated as “Instruction Follower”, has a similar map without the route (see Figure 3.3(b)). The speakers are told that their goal is to reproduce the Instruction Giver’s route on the Instruction Follower’s map. However the maps are not identical (compare Figure 3.3(a) and Figure 3.3(b)) and the speakers are informed about this explicitly at the beginning of their session, but they do not know how exactly the two maps differ.



(a) Map of the Instruction Giver



(b) Map of the Instruction Follower

Figure 3.3: Maps used in the original map task

In our version of the Map Task, the tehnical setup is different than the original one introduced by [6]. We designed two separate cases. In the first, the subject interacted with a computer, but actually the experimenter sim-

ulated the responses of the system. This part of the Map Task Corpus was collected through the Wizard-of-Oz (WOZ) approach.
In the second case, the subject was told they would be interacting directly with another person. It's important to understand that subjects did not go through both cases. Rather, a number of subjects completed the interaction described in the first case and another batch of subjects completed the interaction described in the second case. Section 3.2.2 and Section 3.2.3 detail the differences and similarities concerning the data collection, task description and participants, for both cases.

In summary, our Map Task Corpus consists of the entire collection of videos recorded in both cases. The only tangent point between the two cases is the participant's goal, and that is to guide the computer or the human along the path and reach the end point on the map.

3.2.1 Recording Procedure

The recording procedure described in this section is common to both the Human-Computer Interaction case as to the Human-Human Interaction.
During the quick break that followed the self-presentation session, the experimenter informed the subject of the following steps and tasks. Nothing changed from the self-introduction session setup, meaning that the two participants, namely the experimenter and the subject, were still in separate rooms and communicating via Skype.

A particular characteristic of our data, are the four different behaviors (collaboration levels) that we introduced in the task. Specifically, in order to provoke possible behavioral reactions in the subjects, eliciting their personality traits, the experimenter exhibited four specific behaviors (differently collaborative) during the interactions along the experimental sessions. The four collaboration levels (CL) and the order in which they were played during the interaction are:

- Fully collaborative (CL1): the experimenter was fully complying with the subjects requests (e.g., “perfetto, ho capito, va bene“ lit. “perfect, I understand, all right“ uttered with an enthusiastic and trusting voice)

- Two intermediate collaborative (CL2, CL3): the experimenter used a combination of the two extremes in a neutral, unoffensive tone, using phrases such as ”Che cosa ci devo fare, non é colpa mia” - lit. “What should I do, it is not my fault”or ”Puó ripetere per favore? Mi dispiace non é colpa sua, sono io che non riesco a capire“ - lit. ”Can you repeat, please? I am sorry, it is not your faults, it’s me that I cannot understand”.

- Fully non-collaborative (CL4): the experimenter was behaving opposite to CL1, not complying with the subject’s requests up to being aggressive and offensive (e.g., “É meglio che impari un po’ a spiegarsi“ lit. “it’s better that you learn to explain“ uttered with an aggressive voice).

The experimenter interactions went from a behavior fully collaborative (CL1) to a behavior fully non-collaborative (CL4), going through two intermediate collaboration levels (CL2 and CL3). According to our hypotheses, collaborative behaviors elicit the manifestation of traits related to sociability and positive outcomes (e.g., Agreeableness and Extraversion), while non-collaborative behaviors may trigger anxious reactions and the manifestation of related traits (e.g., Emotional Stability/Neuroticism) in the subjects.

3.2.2 Human-Computer Interaction

In this section we discuss the data collected in the human-computer condition. We cover the subject's task, the collecting procedure, and how the four collaboration levels were employed in the data collection. In the following sections we describe the task, presente the experimental procedures for the HCI scenario and also present the statistics about the participants who took part in the Human-Computer Interaction video recordings.

Task Description

In our version of the Map Task, subjects played the role of the Instruction Giver. Similarly to [25], their maps were changed in order to have a landscape orientation for a better visualization on the laptop screen. We also substituted all the drawings with pictures to make the elements on the map more easily recognizable (see Figure 3.4).

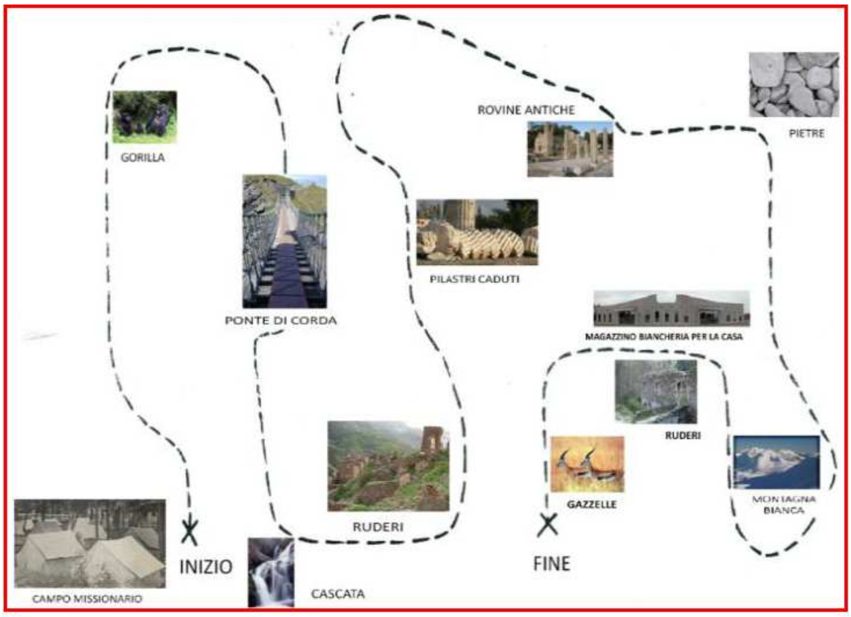


Figure 3.4: The map used by our participants

As in the original version of the task, both the subject and the experimenter had a map with some objects: some of these objects were in the same exact position and marked with the same label, but most of them were in different positions or did not have identical labels. Others were different objects (e.g., a banana tree instead of gorillas close to a palm).

In this condition, the subject was told that they would be interacting with a computer and their task was to guide the computer along a path, marked on the map shown on the computer screen. The machine was actually simulated by the experimenter by means of pre-recorded audio files, that were altered as to resemble a robotic voice. In order to lead the subject to believe they were interacting with a computer, at the beginning of the Map Task session, the experimenter orally simulated a connection to an automatic system by saying "Ora la connetto al sistema e l'esperimento ha inizio" (lit. "Now I am going to connect you to the system and then the experiment will start").

Unlike the original Map Task, where the final goal was to draw on the map of the Instruction Follower the same route reported on the Instruction Giver's map, in our case the Instruction Giver was told to guide the computer from a point to another, according to the route marked on their map. The system to guide was simulated by the experimenter who was playing a set of pre-recorded sentences. Such sentences were modified, to be played back in a more robot-like fashion, in order to give the impression of interacting with a computer. During the interaction, the experimenter could use neutral sentences, distinguished in "positive-neutral" (e.g., "sí", "OK", "va bene" lit. "yes", "OK", "fine") and "negative-neutral" (e.g., "no", "cosa¿', "non ho capito" lit. "no", "what¿', "I did not understand").

All four behaviours, performed by means of four different set of specific sentences, were played along the experimental session in the same order for each subject. The only variable was the length of each interaction modality,

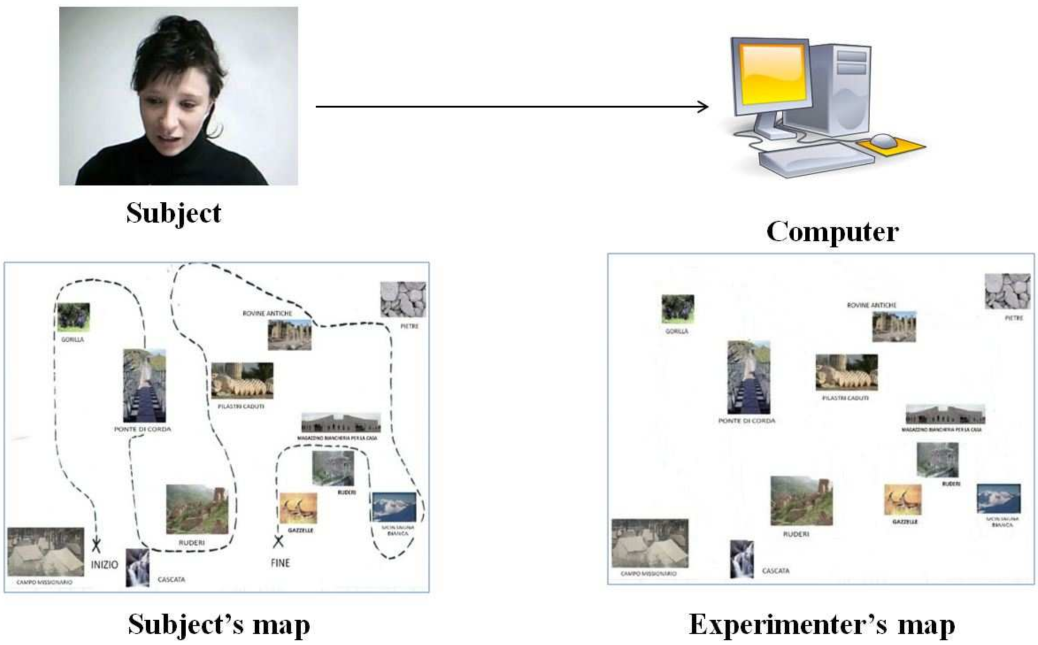


Figure 3.5: Experimental Procedure for HCI

which varied with each subject according to the clarity with which they provided directions. The length of each of the four interaction modalities, varied from 2 to 5 minutes.

Participants

A number of 45 subjects, out of the total of 89 subjects that participated in the self-presentation session, were assigned the task to guide the computer along the path on the map. Because of the poor quality of the collected data during two experimental sessions (repetitive loss of audio and/or video signals), two subjects were left out and thus we remained with a set of 43 subjects.

The distribution of the participants was quite balanced in terms of gender (19 female and 24 male) and age (18 people under 25 and 25 people over the age of 25). The average age the subjects was 31, falling within the adult range (over 25).

Personality Questionnaire and Scores

As previously mentioned, before the start of the experimental session, the participants were asked to fill in the Italian version of the Big Five model personality questionnaire, validated on the Italian language [92]. We computed the new set Big Five scores, for the participants who were part of HCI setting, according to the procedure in [92]. The results (average and standard deviation values) are reproduced in Table 3.3.

Table 3.3: Averages and standard deviations of Big Five Scores on the Maptask Corpus - HCI scenario

	Women		Men	
	<25	≥ 25	<25	≥ 25
Agreeableness	47.5 3.96	50.91 6.3	48.1 9.67	50.57 8.58
Conscienciousness	39.75 9.45	50.09 9.05	38.4 9.37	49.07 8.08
Emotional Stability	35 7.62	36.82 4.14	41.4 6.5	46.07 6.84
Creativity	42.25 4.2	44.73 6.07	48.8 7.38	47.36 8.19
Extraversion	44.88 10.05	46.64 7.3	45.06 9.05	41.36 7.59

Similar to the analysis of the dependence of traits on gender and age, done in Section 3.1, we also looked in this case for any dependency between personality traits and age and gender. After splitting the participants in two classes: people younger than 25 and people 25 or more years old, we examined the data by means of a series of Analysis of Variance (ANOVA). The factorial scores for Extraversion, Agreeableness, Conscientiousness, Emotional Stability and Creativity were the dependent variables, and Age and Gender were used as predictors. The results of the analyses showed that no effects were found for any of the five personality traits. The number We believe the reduced number of instances are not sufficient to uncover significant patterns in the data.

3.2.3 Human-Human Interaction

The information presented in this section serves to highlight how the data collected in the human-human setting is similar and different to the one presented in Section 3.2.2.

Task Description

The procedure is the same as the one for the Human-Computer Interaction, the only difference is that the subjects were told they would be interacting with another person. Their task was to guide this person along the path marked on the map. Also in this version of the Map Task, subjects played the role of the Instruction Giver, the person sitting in the experimenter room and interacting with the subject was the Instruction Follower. The map is the same as the one used in HCI and presented in Figure 3.4.
As in the original version of the task, both the subject and the experimenter had a map with some objects: some of these objects were in the same exact position and marked with the same label, but most of them were in different positions or did not have identical labels. Others were different objects (e.g., a banana tree instead of gorillas close to a palm).
after the subject was introduced to their partner, the experimenter gave the to-go signal.
Unlike the original Map Task, where the final goal was to draw on the map of the Instruction Follower the same route reported on the Instruction Giver's map, in our case the Instruction Giver was told to guide the other person from a point to another, according to the route marked on

their map. During the interaction, the person could use neutral sentences, distinguished in “positive-neutral“ (e.g., “sí“, “OK“, “va bene“ lit. “yes“, “OK“, “fine“) and “negative-neutral“ (e.g., “no“, “cosa¿‘, “non ho capito“ lit. “no“, “what¿‘, “I did not understand“).
All four behaviours, performed by means of four different set of specific sentences, were played along the experimental session in the same order for each subject. The only variable was the length of each interaction modality, which varied with each subject according to the clarity with which they provided directions. The length of each of the four interaction modalities, varied from 3 to 6 minutes.

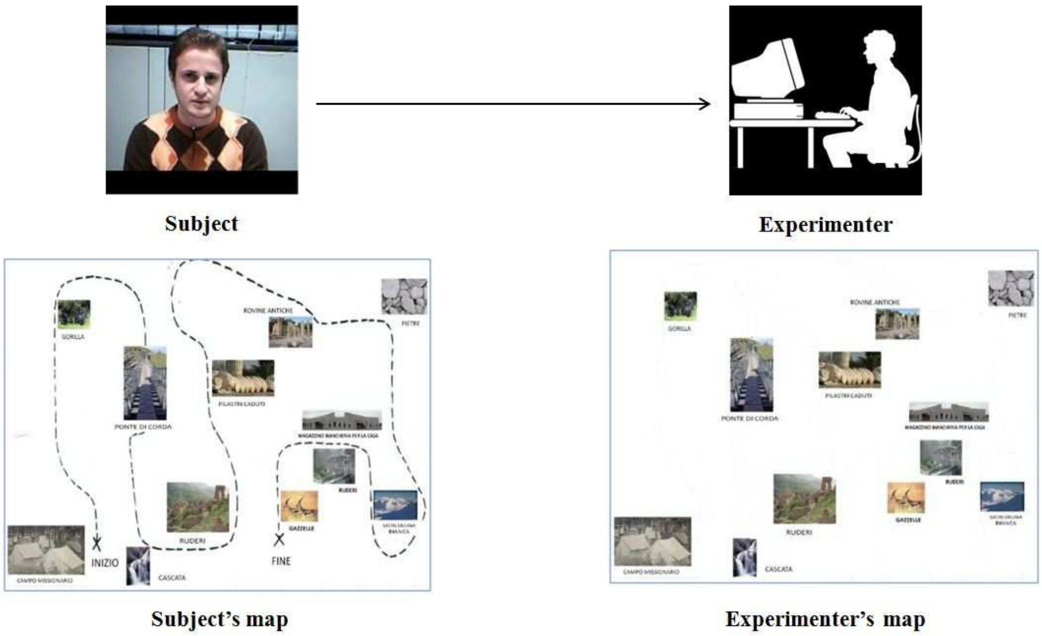


Figure 3.6: Experimental Procedure for HHI

Participants

A number of 31 subjects, out of the total of 89 subjects that participated in the self-presentation session, had to guide the other person along the path on the map.

The distribution of the participants was quite balanced in terms of gender (15 female and 16 male). In total, 11 females were under 25 and 4 were over, while 14 males were under 25 and 2 were over. The average age of the subjects was 25.

Personality Questionnaire and Scores

As previously mentioned, before the start of the experimental session, the participants were asked to fill in the Italian version of the Big Five model personality questionnaire, validated on the Italian language [92]. Taking into consideration only the participants who were recorded in the HHI setting, we computed the the Big Five scores according to the procedure in [92]. The results (average and standard deviation values) are reproduced in Table 3.3.

Table 3.4: Averages and standard deviations of Big Five Scores on the Maptask Corpus - HHI scenario

	Women		Men	
	<25	≥ 25	<25	≥ 25
Agreeableness	47.56 8.46	44.48 7.59	47.98 8.18	51.02 8.80
Conscienciousness	47.11 10.11	40.08 6.66	49.28 10.56	45.87 11.28
Emotional Stability	56.59 7.94	53.56 8.07	51.98 8.63	46.90 13.17
Creativity	47.98 10.25	56.06 11.37	47.45 9.55	54.53 3.41
Extraversion	52.93 5.96	50.81 3.31	44.74 9.09	50.41 1.40

4. Recognizing Personality from Self-Presentation Videos

Social psychology research has shown that personality plays an important role in the way people manage the image they convey of themselves in self-presentations and employment interviews, trying to affect the audience's first impressions and increase the effectiveness of [67].
At the same time, other studies have provided evidence that interviewers perform personality inferences during personnel selection interviews: for instance, Huffcut et al. [55] found that personality traits and social skills were the most frequently measured constructs of potential employees.
Moreover, the type and degree of interview structure seems to moderate this practice. Roth et al. [99] found that the extent to which interviews inadvertently measure an applicant's personality seems to depend on the extent to which interpersonal skills are allowed to play a role throughout the interview process. Nonverbal cues play a major role in this complex process ([46], [94]). DeGroot and Gooty [36] examined the mutual influences between personality attributions, performance ratings and interviewees non-verbal behaviors. Using a structured behavioral interview setting, they found that raters can make personality attributions using only one channel of information (e.g., acoustic, visual, etc.) and that these attributions mediate the relationships between the interviewees nonverbal cues and performance ratings. Moreover, Conscientiousness attributions

can explain the relationship between visual cues and interview ratings. Extraversion attributions mediate the relationship between vocal cues and interview ratings, and Neuroticism attributions had a suppressing effect for both visual and vocal cues. Regarding Extraversion, the results show that the interviewers infer this trait only from the speakers voice characteristics.

Another area of relevance is that of online dating. People who are active on dating websites are fully aware of the importance of their first impression, expressed through their profile or a self-introduction video, on finding a possible partner. When it comes to presenting themselves, online dating service users make a compromise between an accurate self-presentation and how they would actually like to be perceived by others [39]. Regarding misrepresenting themselves, Hall and his colleagues, studied the effect of different factors, such as gender and personality, and found that men are more likely to misrepresent their personal assets, relationship goals, personal interests, while women lie about their weight. As for personality, Agreeableness, Conscientiousness, and Openness to Experience showed consistent correlation with misrepresentation in online dating [48].

Following much psychosocial work on first impressions formation, we exploit the, earlier introduced, concept of thin slices [4] to refer to the short amount of expressive behaviors that we, humans, rely on to produce impressively precise judgments about an individual or a groups properties, such as personality, teaching capabilities, negotiation outcomes, etc.

In this chapter I present our study on automatic personality recognition from the self-presentation collection of videos. The work presented in this chapter brings the following contributions to the existing body of literature on personality recognition. First, we focus on a multimodal approach and second, to our knowledge, the self-presetation scenario has not been used before for the task of automatically recognizing personality.
The sections in this chapter are: Section 4.1 which details the extractation

and processing of visual cues. Section 4.2 presents the audio cues extracted from the speech signal, and also the additional features that were extracted. Section 4.4 presents the machine learning algorithms used for the task of automatic personality assessment. Section 4.5 presents the results of the study and Section 4.6 provides a detailed discussion and interpretation of the results.

4.1 Visual Features

Visual cues play an important role when it comes to behavioral analysis. Gestures, have the power to reveal about their personality style and this is the motivation we are Brebner pointed out that the frequency of hand movements and gestures is positively correlated with Extraversion [21]. Riggio [97] suggested that extraverts have a higher head movement frequency than introverts and change their posture more often. Moreover, they maintain eye-contact for longer [99] and have higher speaking and gestural fluency, while laughing has been associated with high scores in Openness to Experience/Creativity [78]. Many visual cues have been correlated to Agreeableness [43], [18] and Conscientiousness has been correlated with gaze, speaking fluency, speech rate and hand movements ([43], [18]). Their importance in revealing so much about a person, is what stands at the root of our motivation to extract these cues and investigate them. Is what motivation behind our actions.

4.1.1 Manually Annotated Visual Cues

Since visual activity bears discriminative power when it comes to making personality judgements, we manually annotated a series of behavioral and visual cues, indicative of personality traits. The manually annotated descriptors are presented in Table 4.1. For the annotations we used the

ANVIL tool [61].

Table 4.1: Manually annotated visual cues

Cues	Label	Notes
Eye-Gaze	Up	upward directed gaze
	Down	gaze directed down
	Right	gaze directed right
	Left	gaze directed left
	ClosedEyeLid	lids closed for longer than 2 sec
	DesktopCtc	gaze directed towards the desktop
	CamCtc	gaze directed towards the webcam.
Frowning	Yes, No	-
Hand Movement	MovFace	hand(s) move on the face
	MovAir	hand(s) moving in the air
	MovBody	hand(s) move on parts of the body, except face
	StillFace	hand(s) still on face
	StillAir	hand(s) still in the air
	StillBody	hand(s) still on parts of the body, except face.
Head Orientation	Left	head oriented left
	Right	head oriented right
	Down	head oriented down
	Up	head oriented up
	Front	head oriented frontal
	RightSide	head tilted right
	LeftSide	head tilted left
	RightIncl	head half oriented right
	LeftIncl	head half oriented left
Mouth Fidgeting	Smile	smiles
	tongueLips	passing their tongue over their lips
	biteLips	subject biting their lips
	tightLips	subject pressing their lips
	retractLips	move lips by lowering both mouth corners.
Posture	Back	subject leaning back
	Straight	subject in straight posture
	Foward	subject leaning fowards

4.1.2 Computed Visual Features

From the hand-annotated cues presented in Table 4.1 we computed in total 17 visual features. The complete list of visual features and what they represent is presented in Table 4.2.

4.2 Audio Features

The importance of acoustic features, such as pitch and intensity, for personality has been pointed out by many studies, such as ([37], [101]). Furnham [44] discussed how extraverted people are characterized by a particular speaking style: they talk more, louder, faster and have fewer hesitations. Conscientiousness was positively correlated with speaking fluency and speech rate [9]. Mohammadi et al. [81] showed how prosodic features can be used to predict the personality assessments by human experts on a collection of 640 speech samples.
Drawing on this related literature, we derived a set of audio cues, extending their usage to traits for which less evidence is available, e.g., Creativity. Pitch and acoustic intensity were automatically extracted using Praat software [17]. The pitch algorithm is based on the autocorrelation method. Setting the default values for the time step and pitch ceiling (600 Hz) parameters, but lowering the pitch floor from 70 Hz to 50 Hz, the algorithm, uses a time step of 0.015 seconds and a window length of 0.06 seconds. It computes 67 pitch values per second. Intensity is calculated taking into consideration the minimum periodicity frequency if the signal [17]. Table 4.3 provides a detailed account of the acoustic features that were computed automatically, using Praat.

Table 4.2: Visual Features

	Ref	Label	Description
Visual Features	1	avDurationBack	Average duration of leaning back episode
	2	avDurationCam	Average duration of looking into the webcam episodes
	3	avDurationDown	Average duration of looking down episodes
	4	avDurationFrow	Average duration of frownings
	5	avDurationFwd	Average duration of leaning forward episodes
	6	avDurationStr	Average duration of straight posture episodes
	7	freqBack	Rate of leaning backward postures
	8	freqCam	Rate of looking into the webcam events
	9	freqDown	Rate of looking down events
	10	freqForward	Rate of leaning forward postures
	11	freq	Rate of frownings
	12	freqHandMoving	Rate of hand movement events
	13	freqHandStill	Rate of hand still events
	14	freqMouth	Rate of lip moving or biting events
	15	freqSmile	Rate of smiles
	16	freqStraight	Rate of straight postures
	17	NrHeadOrient	Rate head orientation change

Table 4.3: Acoustic Features

	Ref	Label	Description
Acoustic Features	18	MeanPitch	Mean of pitch
	19	MinPitch	Minimum of pitch
	20	MaxPitch	Maximum of pitch
	21	MedPitch	Median of pitch
	22	StDevPitch	Standard deviation of pitch
	23	MeanInt	Mean of Intensity
	24	MinInt	Minimum of intensity
	25	MaxInt	Maximum of intensity
	26	StDevInt	Standard deviation of intensity

4.2.1 Additional Features

It has been shown that dominant people like to speak more and take the floor for a longer period of time [96]. Since these speaking behaviors are encoded in our data and they are relevant to personality, we decided to compute three additional features: the total time of speech, the average duration of voiced segments and the length of the self-presentation. They are presented in Table 4.4.

Table 4.4: Additional Features

	Ref	Label	Description
Additional Features	27	avTimeVoiced	Average duration of voiced segments
	28	TimeVoiced	Portion of self-presentation session taken by speech
	29	videoLength	Total length of self-presentation session

4.3 Feature Analysis

The (linear) relationships between our features and the Big Five personality traits were analyzed by means of a number of backward linear regression analyses, one per each trait, with all our features as predictors.
Table 4.5 reports for each of the final models the retained predictors and the portion of dependent variable variance explained. As can be seen, the linear models only account for a small portion of variance.

Table 4.5: Retained predictors and portion of dependent variable variance explained

	Retained predictors	R^2
Agreeableness	16, 20, 29	.127
Conscienciousness	15, 17, 18, 23, 27, 28	.188
Creativity	10, 12	.107
Emotional Stability	5, 10, 12, 13, 23	.148
Extraversion	6, 11, 15, 27	.172

Table 4.6 reports the partial correlations between the retained predictors and the various traits. Agreeable people tend to have a straight posture more often, to have a lower maximum pitch and produce longer self-presentations. More conscientious people smile more frequently and use a longer portion of their presentation for speaking than less conscientious individuals; but they move their head around less, tend to exhibit a lower average pitch and lower minimal vocal energy, and produce shorter (on average) voiced segments. More creative people lean towards the camera more often than less creative ones but gesticulate less. Higher emotional stability is associated with a greater number of rather short leaning forward events, lower amount of gesticulation and lower vocal intensity. Finally, extraverts produce longer voiced segments, but smile and frown less frequently and maintain a stright posture for a shorter time the straight

posture in front of the camera.

Table 4.6: Partial correlations between the retained predictors and the personality traits; see Table 4.2, Table 4.3 and Table 4.4 for the features' reference numbers.

Feature	16	20	29	-	-	-
Agreeableness	.303	-.216	.217	-	-	-
Feature	15	17	18	23	27	28
Conscientiousness	.203	.208	-.337	-.207	-.183	.274
Feature	10	12	-	-	-	-
Creativity	.321	-.195	-	-	-	-
Feature	5	10	12	13	23	-
Emotional Stability	-.280	.290	-.258	.287	-.195	-
Feature	6	11	15	27	-	-
Emotional Stability	-.259	-.184	-.265	.240	-	-

4.4 Classification Experiments

4.4.1 Feature Ranking

A known problem in classification tasks is to find strategies to reduce the dimensionality of the feature space in order to avoid over-fitting. In our experiment, we applied the Weka implementation of the Support Vector Machine Recursive Feature Elimination (SVM-RFE) algorithm (called Support Vector Machine attribute evaluation method in Weka) in order to evaluate the importance of a feature. This algorithm was introduced by Guyon et al. [47] in a cancer classification problem with the goal of finding a subset of features which maximize the performance of the classifier. The SVM-RFE algorithm creates a weight vector, where a weight is assigned to each feature. The weight vector is used to determine the least important feature, defined as the one with the smallest weight. At each iteration, the least important feature is removed and the procedure is repeated on the reduced feature set. This method is used with a Ranker search method and the features are ranked according to the square of the weights assigned to them. Hence, the first feature is the most relevant for the classification task at hand and the last feature of the least relevant one.

4.4.2 Automatic Classification

For the sake of our classification experiments, all personality traits scores were quantized (Low/High) along their median values (Agreeableness = 50, Conscientiousness = 44, Creativity = 46, Emotional Stability = 39, and Extraversion = 45).
Three machine learning algorithms, namely Naïve Bayes, SVM with linear kernel and SVM with Radial Basis Function kernel, were used in 5 binary classification tasks, one per personality trait. The bound-constrained SVM classification algorithm was used for the two SVM classifiers. The cost parameter C and the RBF kernel parameter C were estimated through an inner leave-one-out cross validation on the training set of the first fold using the first 88 subjects for the parameter estimation. The best performing values of the parameters were then kept fixed for the outer-cross validation.
Three machine learning algorithms, namely Naïve Bayes, SVM with linear kernel and SVM with Radial Basis Function kernel, were used in 5 binary classification tasks, one per personality trait. The bound-constrained SVM classification algorithm was used for the two SVM classifiers. The cost parameter C and the RBF kernel parameter γ were estimated through an inner leave-one- out cross validation on the training set of the first fold using the first 88 subjects for the parameter estimation. The best performing values of the parameters were then kept fixed for the outer-cross validation.
For each classifier and for each trait, we executed 29 classification runs,

each exploiting a subset of the features aggregated according to the ranking provided by the SVM-RFE algorithm. We started from the single feature experiment using the first ranked feature, then executed the 2-feature experiment with the first two ranked features, and so on. The leave-one-out cross-validation strategy was employed. Hence, 89 models for each personality trait were trained on 88-subject subsets, evaluating them against the remaining ones and finally averaging the results.

4.5 Classification Results

Table 4.7 reports, for each classifier and for each trait, the feature combination producing the highest accuracy value. All the accuracy values reported are statistically significant, according to binomial tests that compared the observed accuracy to that of the baseline classifier that exploits the observed frequencies of the two classes. The significance was set at $p<=0.01$ and the resulting threshold values for accuracy were 0.618 for Agreeableness, Conscientiousness and Emotional Stability, and 0.629 for Creativity and Extraversion. The highest accuracy values, greater or equal to 70%, were obtained on Conscientiousness (SVM-RBF), Emotional Stability (SVM-RBF and SVM-Lin) and Extraversion (SVM-RBF and Bayes). To further characterize the predictive power of our features and the behavior of the classifiers, we investigated how they worked on the two classes each trait was split into. We started from the confusion matrix of the conditions in Table 4.7 and compared the hits for the High and Low classes with those expected from the baseline classifier. The comparison was conducted by means of Pearson residuals, standardized scores - they are N(0, 1) - that measure the difference between observed and expected outcomes. On hits, the absolute value of a Pearson residual measures how much the classifier performs better (positive sign) or worse (negative sign) than the baseline in terms of recall. For errors, the reverse is true. Here we focus on hits.

Finally, we took advantage of the *N(0,1)* distribution of the Pearson residual and fixed a threshold of ±3SD for the statistical significance of the difference between observed and expected hits. Table 4.8 reports the results.

Table 4.7: Accuracy (Acc), recall on the Low class (Rec1) and on the High class (Rec2) for the best condition

Classifier	SVM_RBF			SVM_Lin			Bayes		
	Acc	Rec1	Rec2	Acc	Rec1	Rec2	Acc	Rec1	Rec2
Agreeableness	65.16	50	80	65.16	50	80	64.04	34.09	93.33
Features	8, 22, 6, 15, 29, 10			8, 22, 6, 15, 29, 10			8, 22, 6, 15, 29, 10, 16		
Conscientiousness	**73.03**	73.91	72.09	68.53	76.09	60.47	-	-	-
Features	19, 22, 27, 6, 14, 24, 20, 17, 25			19, 22, 27, 6, 14, 24, 20, 17, 25, 12, 9, 16			-		
Creativity	64.04	72	53.85	66.29	76	53.85	64.04	68	58.97
Features	16, 8, 15, 26, 12, 7, 6, 4, 24, 20, 14, 13, 2, 5, 1,			16, 8, 15, 26			16, 8, 15, 26, 12, 7, 6, 4, 24, 20, 14, 13, 2, 5, 1,		
Emotional Stability	**76.40**	80.44	72.09	**75.28**	82.61	67.44	64.04	45.65	83.72
Features	25, 22, 20, 13, 12, 10, 16, 9, 2, 19, 8			25, 22, 20, 13, 12, 10, 16, 9, 2, 19			25, 22, 20, 13, 12, 10, 16, 9, 2		
Extraversion	**70.78**	76.60	64.29	-	-	-	**69.66**	74.47	64.29
Features	24			-			24		

4.6 Discussion

SVM-RBF was the only classifier yielding balanced performances on at least two traits: Conscientiousness and Emotional Stability. In all the other cases, good performances arose only from either the Low or the High

Table 4.8: Pearson residual for the condition of Tables 4.2, 4.3, 4.4. Only residuals greater or equal in absolute value than 3 are reported.

-	SVM_RBF		SVM_Lin		Bayes	
-	Low	High	Low	High	Low	High
Agreeableness	-	4.25	-	3.95	-	5.74
Conscienciousness	3.02	3.12	3.31	-	-	-
Creativity	-	-	-	-	-	-
Emotional Stability	3.90	3.12	4.20	-	-	4.65
Extraversion	3.27	-	-	-	2.97	-

class. For instance, with Agreeableness good performances are limited to the High class whereas with Extraversion they are limited to the Low class. For Creativity, no classifier yielded significant performances on any class.

It is not by chance that Conscientiousness and Emotional Stability are the traits that yielded the best results, both in terms of accuracy values and of the level of balance between the recall values for the Low and High classes. Conscientiousness, in fact, relates to the individuals capacities for behavioral and cognitive control. Conscientious individuals are described as responsible, attentive, careful, persistent, orderly, and organised; those low on this trait are irresponsible, unreliable, careless, and distractible. High conscientiousness has been connected to positive engagement within task-related behavior [11]. Apparently, the request of introducing themselves in front of a monitor, with a camera and microphone on, activated our subjects Conscientiousness dispositions, doing so for both those high and those low in this trait. Those dispositions, in turn, affected some of the considered behaviors, including: the dynamics of pitch (minimal and maximal pitch, and pitch range as captured by StDevPitch feature) and of voice energy (minimal and maximal intensity) on the acoustic side; posture (average duration of episodes of sitting straight in front of the monitor) and head movements and lip-related events on the visual side.

Emotional stability and its counterpart, neuroticism, concern the susceptibility to negative emotions, which we think might be elicited by the task of introducing oneself in front of a computer screen while being audio-video recorded. Emotional Stability/Neuroticism include both anxious and irritable distress. The former is inner-focused and includes dispositions to anxiety, sadness and insecurity. Irritable distress, in turn, involves outer-directed hostility, anger, frustration and irritation. We can easily figure out that the self-introduction task activates one of those two sets of dispositions, depending of the persons internal constitution. The behavioral signs that proved effective concerned pitch dynamics (maximum and minimum pitch, and pitch standard deviation), maximum voice intensity, and several visual features: dynamics of hand movements, posture dynamics (straight and forward position), camera fixation and camera aversion, and hand fidgeting. According to this line of explanation, both the request of executing a task and the nature of the task (introducing oneself) activated dispositions connected to Conscientiousness and Emotional Stability. We think that a similar rationale can be given to account for the performances with other traits.

A good accuracy level was obtained with Extraversion too, but this result could be due to the good performance with introverts. One might suggest that the inbalance might be (at least partially) due to the fact that the situation/task was not appropriate for fully activating Extraversion-related dispositions, especially with extraverts. It has been argued, in fact, that the core of Extraversion lies in the tendency to behave in a way so as to engage, attract and enjoy social attention, i.e., extraverts invest time and energy in activities that attract the attention of others ([10], [71]). One possible consequence of this view is that Extraversion-related dispositions are activated to a lesser extent in situation like the one we are considering here where there are no others from who to attract attention, and that this should affect extraverts' behavior especially. Notice, finally, that only one feature (minimal vocal intensity, MinInt) is used to obtain the accuracy value for Extraversion when using the SVM with an RBF kernel.

Agreeableness does not reach the same levels of accuracy as Conscientiousness and Emotional Stability. However, the recall for the High class, both measured in absolute term (see Table 4.7) and through Pearson residuals (see Table 4.8) is very high: in absolute terms, it reaches 80% with SVM-RBF and SVM-Lin and 93% with Naïve Bayes. This trait includes a number of dispositions that foster congenial social behavior: generosity, consideration, cooperation, willingness to accommodate others wishes, etc. Agreeable people do not aim to attract social attention like extraverts, but to please others. We believe that, again, the key to understanding the performance of our features and classifiers with this trait is in the nature of the situation: it activates a pleasing attitude that somehow masks non-agreeable dispositions (being aggressive, rude, manipulative, etc.).

Finally, the dispositions linked to Creativity seemed to be uniformly activated to much a lesser extent by the task of introducing oneself.

5. Human-Computer Interaction and Personality

Humans have the tendency to understand and explain other humans' behavior in terms of stable properties on the basis of observations of everyday behavior. In this sense, the attribution of a personality and its usage to infer about others is a fundamental property of our naïve psychology.
Scientific psychology has maintained the importance of personality as a higher-level abstraction, encompassing sets of stable dispositions towards action and towards belief and attitude formation. The concept of personality is commonly used to explain human behavior in several domains: clinical and social psychology, educational studies and so on. Most of the existing models of personality are based on traits, meant as higher level abstraction derived on the basis of factorial studies [11]

Bickmore and Picard [15], as well as Nass and Brave [87], argued that matching users' personality increases the efficiency of the interaction between users and relational agents (computational artifacts designed to establish and maintain long-term social-emotional relationships) and make the interaction more natural for the user. Nakajima et. al, [84] have argued that in the context of a human-machine collaboration system, the system's personality is: (a) perceived by the user, and (b) influencing the user's behavior. Also, users adapt their behavior to fit the system with which they are interacting. This adaptation reflects beliefs about the systems capability, and not its actual behavior [91].

In the field of Human-Computer Interaction (HCI), personality is a key factor [72]. Nass and colleagues showed that participants classified as submissive significantly preferred interacting with a computer that exhibits a submissive behavior, while dominant participants prefer to interact with a more dominant computer [86]. Their extended work with computer-synthesized voices showed that extraverts related better to an extraverted voice, in terms of liking, trusting and following its instructions, while introverts preferred an introverted artificially generated voice signal [85].

The notion of personality has been used to improve the believability of life-like characters. The basic idea is that a virtual agent can appear more realistic if the engine that plans its actions can condition the behavior according to a given personality scheme (see for example [7] and [35]). When the goal is to develop an interactive robot, it is important to invest time and effort to endowe it with properties that would make it acceptable by human users. This contributes to a successful interaction, and attention must be paid to both the appearance and the personality of the robot ([50], [56]), even if only the robot's face benefits from personality generation.

Within the research area of socially assistive robotics [40], Ţăpuş et al. [107] tackled the issue of matching the assistant robot's personality to the personality of post-stroke patients who were involved in rehabilitation therapy. However, they focused only on Extraversion/Introversion trait.
[56] Recently, several works have explored the automatic analysis of personality ([14], [70], [93], [80]) in different scenarios, often targeting the all of the Big Five (Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Creativity) or just a subset.

In this section, we will detail the methodology and findings of the attempt at proceeding one step further and investigate to which degree it is possible to automatically predict all five traits of the Big Five model in a scenario that resembles one possible vision of HCI: a human interacting with a machine.

To date, few works have addressed the issue of recognizing all the Big Five personality traits in a human-machine interaction context. In particular, to our knowledge, no work has yet addressed the issue of personality recognition in the assistive human-machine interaction context, under the influence of different levels of collaboration between the two interactants.

5.1 Visual features

Based on the same rationale presented in Chapter 4, Section 4.1 we have extracted a series of visual features from the Human-Computer Interaction data. No manual annotations were performed since all visual features have been extracted automatically, using a software that is able to capture ample as well as more subtle movements. More exactly, five visual activity cues were extracted using the approach described in [58]. For each frame where the participant is visible in the close-up view, the average motion vector magnitude (MVM) and the residual coding bitrate (RCB) over all the estimated blocks and skin blocks are computed and used as a measure of individual visual activity. Motion vectors are generated from motion compensation during the video encoding; for each source block that is encoded in a predictive fashion, its motion vectors indicate which predictor block from the reference frame is to be used. A predictor block is highly correlated with the source block and hence similar to the block to be encoded. Therefore, motion vectors are usually a good approximation of optical flow, which in turn is a proxy for the underlying motion of objects in the video [27].

After motion compensation, the DCT (Discrete Cosine Transform) coefficients of the residual signal, which is the difference between the block to be encoded and its prediction from the reference frame, are quantized and entropy-coded. The residual coding bitrate is the number of bits used to encode this transformed residual signal.

Motion vector magnitude and residual coding bitrate capture different kinds of information: while the motion vectors capture the rigid body motion like translation, residual coding bitrate attempts to capture the non-rigid motion. In order to compare the motion vector magnitudes and the residual coding bitrate in a meaningful way, the quantities were normalized [109]. Following [58], we also used a fifth measure: the combined estimate of the visual activity, namely the average of visual activity from motion vector magnitude and from residual coding over the estimated skin blocks. In this way, both rigid and non-rigid local motions can be approximated. All the above mentioned features were used to detect dominant people in a meeting scenario, based on the idea that dominant people tend to move more [58]. Table 5.1 presents all the visual features that we have extracted in an automatic manner.

5.2 Acoustic Features

In our corpus audio cues were extracted in two different ways. On the one hand, pitch, intensity and their mean, maximum, minimum, median and standard deviation were extracted automatically using the audio processing tool, Praat [17]. On the other hand, although manual annotations of dialogue segments can be very tedious and time consuming, we opted to manually label the data in order to enrich the corpus annotations.

Table 5.1: Visual Features

	Ref	Label	Description
Visual Features	1	MVM_All	Motion Vector Magnitude computed over all blocks in the frame
	2	MVM_Skin	Motion Vector Magnitude computed over the skin blocks in the frame
	3	RCB_All	Residual Coding Bitrate computed on all the blocks
	4	RCB_Skin	Residual Coding Bitrate computed over the skin blocks in the frame
	5	MVMRCB_All_Av	Average between the Motion Vector Magnitude and Residual Coding Bitrate computed on all the blocks in the frame
	6	MVMRCB_Skin_Av	Average between the Motion Vector Magnitude and Residual Coding Bitrate computed on the skin blocks in the frame

5.2.1 Speaker Diarization System

When extracting acoustic features related to the dialogue between the subject and the machine, we came across the speaker diarization problem, namely the process of partitioning an input audio stream into homogeneous segments according to the speaker's identity. However, since we know the number of speakers involved in the interaction (always two) and their identity, we have approached the problem from a speaker recognition perspective. We used Gaussian mixture models (GMMs), a background model and maximum a posteriori (MAP) adaptation for updating means of the Gaussian distribution in the speaker models. The background model is constructed using a speech corpus. It is based on two models: one for speech and one for silence. Adaptation process is applied to both of the models using data from the actual corpus that will be annotated (the Map Task).

In our study, we constructed the model for the experimenter by applying MAP adaptation by means of the HTK toolkit [110]. For each subject, a single speech model has been constructed by adapting the background model using speech samples collected from each of them. According to the method, we then applied Viterbi decoding to all the videos for the given subjects, again by the HTK toolkit. In this way, we obtained the automatically annotated videos with the corresponding speech label (for both the subject and the experimenter) and silence label. Finally, the annotations were converted to an ANVIL - compliant format, to make them visible using this annotation tool.

5.2.2 Acoustic Features from Automatic Annotations

Although the quality of the automatic annotations was acceptable in quality, in order to fully take advantage of the importance of each feature component, that would be extracted from the annotations, it was necessary to post-edit them. This was done by manually correcting the label of the erroneous speech segment label. From the final version of manual annotations we derived the following set of features, that capture dialogue dynamics as well as individual speaking behavior:

- **Turns**: computed over the entire dialogue, they refer to the number of turns of the subject. To extract this feature, we took into consideration the number of speech segments labeled with "Subject".

- **Long Turns**: computed over the entire dialogue, they refer to the number of the subjects speech segments that are longer than 2 seconds. During the dialogues we observed events such as back-channeling, coughing or sounds irrelevant to the dialogue. So as not to include these, we decided to discard turns shorter than 2 seconds. A similar approach was proven to be effective in [58] to characterize dominant behavior.

- **Total Speaking Duration of the Subject**: captures the total amount of time in which the subject is vocally active. This feature sums the duration of all speech segments with the "Subject" label.

- **Total Speaking Duration of the Experimenter**: refers to the total amount of time in which the experimenter is speaking. This feature sums the duration of all speech segments with the "Experimenter" label.

- **Total Duration of the Overlapped Speech**: captures the total amount of time where the subject and experimenters speech overlap.

- **Total Duration of the Silence**: For each video, we have computed the total amount of time, in which neither the subject nor the experimenter is active.

5.2.3 Acoustic Features Extracted Automatically

Focusing only on the subjects' speech segments, pitch and acoustic intensity were automatically extracted using Praat [17]. The start and end time of each segment has been taken from the post-edited annotations. The pitch algorithm is based on the autocorrelation method. By setting the default values for the time step and pitch ceiling (600 Hz) parameters, but lowering the pitch floor from 70 Hz to 50 Hz, the algorithm uses a time step of 0.015 seconds and a window length of 0.06 seconds. It computes 67 pitch values per second. Intensity is calculated taking into consideration the minimum periodicity frequency in the signal. The maximum, minimum, average, median and standard deviation were computed for pitch, as well as for intensity. The acoustic features that were extracted with Praat and the ones derived from the automatic annotations are summarized in Table 5.2.

Table 5.2: Acoustic Features

	Ref	Label	Description
Acoustic Features	7	DurSpeechExp	Duration of experimenter's speech
	8	DurSpeechSubj	Duration of subject's speech
	9	DurOverlap	Duration of the overlapping speech between the subject and the experimenter
	10	DurSilence	Duration of silence
	11	MeanPitch	Mean of subject's pitch
	12	MaxPitch	Maximum of subject's pitch
	13	MinPitch	Minimum of subject's pitch
	14	MedianPitch	Median of subject's pitch
	15	StDevPitch	Standard deviation of the subject's pitch
	16	MeanInt	Mean of the subject's speech intensity
	17	MaxInt	Maximum of the subject's speech intensity
	18	MinInt	Minimum of the subject's speech intensity
	19	StDevInt	Standard deviation of the subject's speech intensity

5.2.4 Additional Features

The first two features listed above: the total number of turns between the computer and the subject and the total number of turns between the computer and the subject, that lasts longer than 2 seconds, are ment to capture the contributions of the user's speaking behavior to the dialogue. We labeled them as "additional features" and are presented in Table 5.3.

Table 5.3: Additional Features

	Ref	Label	Description
Additional Features	20	SubjTurns	Number of speaking turns
	21	SubjLongTurns	Number of speaking turns longer than 2 seconds

5.3 Feature Analysis

Same as in Chapter 4, we investigated the linear relationship between the extracted features and the five personality traits. Again, yhe analysis was performed by means of a number of backward linear regression analyses, one per each trait and per each collaboration level, with all features as predictors.

Table 5.4 reports for each of the final models the retained predictors and the portion of dependent variable variance explained. As can be seen, some of the linear models account only for a small portion of variance, while others account for a larger variance.

Table 5.7 (end of this Chapter, page 72) reports the partial correlations between the retained predictors and the various traits, in each of the 4 collaboration levels. In the fully collaborative level, people who move more score higher on the Agreeableness trait. An increased speaking activity and movement is also perceived in people who score higher in Conscientious-

Table 5.4: Retained predictors and portion of dependent variable variance explained.

Condition	Trait	Retained Predictors	R^2
CL1	Agreableness	10, 3, 4, 6	.256
	Conscientiousness	8, 9, 4, 6	.223
	Creativity	8, 10, 11, 14, 6	.188
	Emot. Stability	8, 13, 14, 18, 19, 21	.195
	Extraversion	8, 9, 3, 21	.370
CL2	Agreableness	8, 9, 11, 12, 15, 19, 4, 20, 21, 5, 6	.527
	Conscientiousness	16, 17, 18, 4, 6	.101
	Creativity	10, 11	.164
	Emot. Stability	11, 14, 17, 20, 21	.342
	Extraversion	8, 13, 11, 14, 15, 16, 20	.378
CL3	Agreableness	8, 15, 1, 2, 3, 6	.265
	Conscientiousness	16, 18, 19, 4, 6	.268
	Creativity	7, 8, 11, 1	.257
	Emot. Stability	8, 9, 11, 14, 2, 3, 4	.252
	Extraversion	9, 11, 13, 14, 15, 2, 3, 20, 21, 6	.408
CL4	Agreableness	8, 9, 15, 3	.242
	Conscientiousness	16, 18, 19, 4, 6	.114
	Creativity	7, 14, 16, 17, 18	.196
	Emot. Stability	10, 11, 14	.353
	Extraversion	7, 11, 18, 19, 20, 21	.309

ness and Creativity. A longer speaking duration, minimum voice pitch and intensity are positively correlated with higher scores in Emotional Stability. People who move more and grab the floor more often score higher on Extraversion.

In the two intermediate collaborative levels of interaction, those who move, speak longer, with a higher pitch tend do score higher on Agreeableness, while a lower intensity of the speech signal, together with an increased rate of motion is correlated with Conscientiousness. In this setting, emotionally stable people display an average pitch, their voice intensity is higher and produce longer voiced segments. Extraverts are characterized by the tendency to overlap their speech with that of the computer, but keeping their pitch and intensity close to average values.

In the fully non-collaborative interaction, a longer duration of the speech is still seen as an indicator for Agreeableness. Those who keep a minimum level of the speech intensity and move more, are seen as more conscientious. Higher creativity is associated with a high acoustic intensity and a longer speaking duration from the system. The absence of speech and an average pitch are perceived in people who score higher in Emotional Stability. Finally, also in this type of interaction, longer speaking turns are produced by extraverts.

5.4 Classification Experiments

5.4.1 Feature Selection

Due to time and space constraints, we must restrict the feature space. In our experiments, we applied the Weka implementation of the Classifier Subset Evaluator that evaluates subsets of features using a classifier (in our case, Support Vector Machine with linear kernel) to estimate the 'merit' of a set of attributes [62]. The evaluation was performed on the datasets generated by an inner leave-one-subject-out cross validation. This method was used jointly with a Best First search strategy, that searches the space of subsets of features by 'greedy hill-climbing' augmented with a backtracking facility. We used the forward direction of the search that begins with the empty set of features. In Table 5.5, we report the features selected for the different traits using the linear SVM subset evaluator. These features were used for running the classification tasks in the following subsection.

Table 5.5: Features selected using SVM - Classifier Subset Evaluator. Features' numbers refer to Table 5.1, Table 5.2 and Table 5.3

Cond.	Personality Trait				
	Extraversion	Agreableness	Conscientiousness	Emotional Stability	Creativity
CL1	1, 9, 16, 20	6, 7, 12, 14	1, 2, 6, 7, 9, 13, 14, 18	6, 7, 11	2, 7, 19
CL2	1, 10, 16, 18, 21	1, 4, 9, 10, 14	1, 21	5, 6, 11, 18	3
CL3	14	1, 3, 4, 5, 8, 9, 13, 14, 19, 20	1, 3, 4 , 20	5	1, 2, 7, 16, 17
CL4	2, 10, 17, 20	1, 2, 3, 5, 7, 8, 12, 18	1, 4, 5	5, 7	2, 11, 18, 19

5.4.2 Automatic Classification

For our classification experiments, all personality traits' scores were quantized (Low/High) along their median values (Agreeableness = 50.45, Conscientiousness = 50.72, Creativity = 48.13, Emotional Stability = 51.69, and Extraversion = 50.25).

Classification was performed by means of Support Vector Machines (SVMs) with linear kernel. In particular, the bound-constrained SVM classification algorithm was used for the SVM classifier. The cost parameter C was estimated trough an inner leave-one-subject-out cross validation on the training set of the first fold and were then kept fixed for the outer-cross validation. We designed 5 binary classification tasks, one per personality trait. Each binary classification task was performed for each interaction modality, CL1 - CL4. The total number of experimental conditions was 20. For each classifier, for each trait, and for each collaboration level of the interacting machine, only the features retained in the feature selection step (see Section 5.4.1) were used. The leave-one-out cross-validation strategy was employed. Hence, 43 models for each personality trait were trained on 42-subject subsets, evaluating them against the remaining ones and finally averaging the results. For comparison, we ran the same experiments described above by using SVMs with Radial Basis Function (RBF) kernels and Naïve Bayes. Since the classification using SVMs with RBF kernels and Naïve Bayes gave results not significantly different to those obtained by SVM with linear kernel algorithm, we only report the significant ones.

5.5 Classification Results

Table 5.6 summarizes the accuracy values for each trait and for each level of collaboration (CL). The accuracy values highlighted in bold are statistically significant, according to binomial tests that compared the observed accuracy to that of the baseline classifier that exploits the observed frequencies of the two classes. The significance was set at p$\leq$.01 and the resulting threshold values for accuracy were 0.67 for all the traits.

Table 5.6: Accuracy (%) obtained using SVM with linear kernel

	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Creativity
CL1	62,79	65,12	51.16	**81.30**	48,43
CL2	**81.30**	**69,77**	**69,76**	**74,41**	58,13
CL3	**74,41**	67,44	62,79	**74,41**	58,13
CL4	**72.09**	41,86	60,46	**74,41**	60,46

5.6 Discussion

Extraversion and Emotional Stability are the only traits yielding, in an almost consistent way, results significantly better than the baseline. Emotional stability and its counterpart, Neuroticism, concern the susceptibility to negative emotions, which we think are elicited by the behaviors of the system in the dialogue. All four interaction modalities (CL1 - CL4) seem to fit for the study of the manifestation of this trait. Emotional Stability/Neuroticism includes both anxious and irritable distress. The former is inward-focused and includes dispositions to anxiety, sadness and insecurity. Irritable distress, in turn, involves outward-directed hostility, anger, frustration and irritation. We can easily see that trying to accomplish a given task, in a collaborative or non-collaborative setting, activates one of these two sets of dispositions, depending of the persons internal constitution. The behavioral signs that proved effective were: (i) pitch in all the different conditions given by the machine's role; (ii) intensity (maximum of subject's intensity) while the computer was interacting in fully collaborative way (CL1); and (iii) body motion activity while the computer was interacting in a fully non-collaborative way (CL4).

A good accuracy level was also obtained with Extraversion. The core of Extraversion lies in the tendency to behave in a way so as to engage, attract and enjoy social attention, i.e., extraverts invest time and energy in activities that attract the attention of others ([5], [25]). One possible consequence of this view can be the explanation as to why Extraversion-related disposition are activated to a greater extent in situations like the ones we are considering here. In particular, the effective behavioral signs were: (i) intensity of subjects speech signal while the computer was interacting in fully non-collaborative way (CL4) and in one of the two intermediate modalities (CL2); (ii) body motion activity; and (iii) long speaking turns in case of CL2 intermediate modality.

Agreeableness and Conscientiousness obtained a result slightly above the threshold only in the condition CL2. In particular, for Agreeableness we could hypothesize that there is no coincidence that this trait is better recognized in the CL2 setting. The key might lie in the kind part of the experimenter while CL3 and CL4 might have no bearing: it activates a pleasing attitude in the subject.

We also tested if the performance differences among the different conditions given by different interaction modalities performed by the machine were significant. We did not find any statistically significant difference with the exception of CL1 that gives significantly worse performance for the Extraversion trait than the ones we obtained in the other three conditions. This results call for further investigations.

Finally, we found an interesting trend for the fully collaborative modality (CL1) to facilitate better performances in the recognition of Emotional Stability among the different traits and for CL2 modality to facilitate better performance for classifying Extraversion trait. In a nutshell, the four different collaborative levels exhibited throughout the interaction, have elicited in our participants behaviors typical for the different five personality traits. This draws the attention towards the fact that certain typical behaviors are better observable in certain environments and under certain circumstances.

Table 5.7: Partial correlations between the retained predictors and traits; see Table 5.1, 5.2 and 5.3 for the features' reference numbers. Agre. - Agreeableness, Consc. - Conscientiousness, Crea. - Creativity, Em. Sta. - Emotional Stability, Extr. - Extraversion.

CL1	Features	10	3	4	6	-	-	-	-	-	-	-
	Agre.	-.413	-.292	.262	-.242	-	-	-	-	-	-	-
	Features	8	9	4	6	-	-	-	-	-	-	-
	Consc.	-.231	.447	.271	-.312	-	-	-	-	-	-	-
	Features	8	10	11	1	4	6	-	-	-	-	-
	Crea	-.241	-.345	-.355	.395	.340	-.335	-	-	-	-	-
	Features	8	13	14	18	19	21	-	-	-	-	-
	Em. Sta.	.449	.287	-.325	.259	.261	-.358	-	-	-	-	-
	Features	8	9	3	21	-	-	-	-	-	-	-
	Extr.	-.556	-.383	.376	.255	-	-	-	-	-	-	-
CL2	Features	8	9	11	12	15	19	4	20	21	5	6
	Agre.	.407	-.329	-.238	.265	-.451	-.287	.191	.343	-.276	-.182	-.219
	Features	16	17	18	4	6	-	-	-	-	-	-
	Consc.	-.305	.284	.306	.395	-.399	-	-	-	-	-	-
	Features	10	11	-	-	-	-	-	-	-	-	-
	Crea	-.307	-.402	-	-	-	-	-	-	-	-	-
	Features	11	14	17	20	21	-	-	-	-	-	-
	Em. Sta.	.397	-.395	.433	-.482	.373	-	-	-	-	-	-
	Features	8	13	11	14	15	16	20	-	-	-	-
	Extr.	-.246	-.305	.304	.262	.300	-.298	.428	-	-	-	-
CL3	Features	8	15	1	2	3	6	-	-	-	-	-
	Agre.	.281	-.481	-.355	.330	-.289	-.290	-	-	-	-	-
	Features	16	18	19	4	6	-	-	-	-	-	-
	Consc.	-.403	.354	.367	.397	-.447	-	-	-	-	-	-
	Features	7	8	11	1	-	-	-	-	-	-	-
	Crea	-.455	-.282	-.239	.272	-	-	-	-	-	-	-
	Features	8	9	11	14	2	3	4	-	-	-	-
	Em. Sta.	.244	-.245	.293	-.306	-.338	-.280	.422	-	-	-	-
	Features	9	11	13	14	15	2	3	20	21	6	-
	Extr.	.398	-.318	.260	.281	.256	.202	.375	.235	-.319	-.248	-
CL4	Features	8	9	15	3	-	-	-	-	-	-	-
	Agre.	.375	-.245	-.487	-.305	-	-	-	-	-	-	-
	Features	16	18	19	4	6	-	-	-	-	-	-
	Consc.	-.376	.343	.328	.324	-.335	-	-	-	-	-	-
	Features	7	14	16	17	18	-	-	-	-	-	-
	Crea.	.279	-.448	-.274	.409	-.244	-	-	-	-	-	-
	Features	10	11	14	-	-	-	-	-	-	-	-
	Em. Sta.	.387	.387	-.399	-	-	-	-	-	-	-	-
	Features	7	11	18	19	20	21	-	-	-	-	-
	Extr.	.425	-.297	.325	.305	-.226	.312	-	-	-	-	-

6. Human-Human Interaction and Personality

Valid models for personality recognition in human-computer interaction require an accurate characterization of the interaction and the emergent personality traits between two people. One of the long term goals in social and computer science research is to identify and develop models for personality recognition that contribute to the advancement of intelligent systems and strenghten the envisioned course of computers as social actors.

Our work is based on the assumptions that a significant component of personality models in human-computer interaction require a second, typically more well-known model that serves as baseline. It follows, then, that a well determined version of the human-human personality model can serve when it comes to determine a human-computer interaction model, as a learning and also a complementary model. This will help its counterpart, human-computer interaction perosnality models, perform better.

However, this approach requires that the human-human personality model be able to recognize and react to an adult or child's personality unfolded in an interaction. To be able to model human behavior for future intelligent systems, we must see if the same kind of effects that dominate human-human interactions, particularly those that bring important social benefits, are present in human-robot interactions.

Personality, intelligent systems and interaction have been jointly investigated in connection to a very relevant term: the *similarity-attraction* affect. In social psychology the similarity-attraction effect refers to the propensity to gather around and interact with humans that are similar to us on many levels, such as political affilication, views on different life problems, social and economical status. Researchers Paul Ingram and Michael Morris at Columbia University conducted a study whose results showed that famous expression "opposite attract" is not necessarely true. They found that people tend to mix and interact with other people with whom they identify themselves with, either on the professional or personal level [57]. This idea has also found supporters in the computational domain. A study by Nass et al., [86] showed that submissive people prefer to interact with a computer that exhibits a submissive behavior, while dominant participants prefer to interact with a more dominant computer. Their extended work with computer-synthesized voices showed that extraverts related better to an extraverted voice, in terms of liking and trust, while introverts preferred a corresponding, introverted voice [85].

These findings are not shared unanimously by other researchers. For instance, Lee and colleagues [68] showed that extravert participants rated the interactive robotic dog as socially attractive when it was behaving like an introvert and that introvert participants prefer when the robot behaves in an extraverted manner. This is the opposite of the similarity-attraction effect in terms of personality and interaction. These insightful and thought provoking results, provide the perfect framework to study the emergence of the five personality traits, in a human-human interaction.

An appropriate application domain would be socially-assistive applications. These applications could compute a model of the personality of the user and, based on additional algorithms, could provide in-context help in forming social connections and building common ground between people, for example, at parties or informal meetings. Roughly speaking, it would be an automatic, socially-aware friend or connection recommender system.

Variations in who is the addressee can be influential on the outcome of the personlaity emergence and recognition. Therefore, it is not just the four levels of collaboration but also the different conditions that can influence the emergence of the five personality traits. It must therefore be looked at if personality inferred in a HCI can benefit from personality models inferred from a human-human interaction. To reach our goals and in order to be consistent with the analysis of the emergence of personality traits in human-computer interaction, we follow the same methodology for processing the data and the same algorithms to extract features and perform the classification tasks.

6.1 Visual Features

The behavior, such as gestures or certain positions, that the user brings into an situation or interaction, wheter intended or not, contribute to the expression of personality. When interacting with a peer, people tend to give away their current affective state and the same happens when the interaction is not face-to-face. People gesture, move around and make facial expressions while speaking on the phone, in a similar way as when they are face-to-face with the adresee. Having a rich source of information and for the sake of consistency and comparability, we decided to extract the same set of visual features from the video recordings, as the ones presented in Section 5.1. Since motion vectors are a good approximation of optical flow, which in turn is an estimation for the underlying motion of objects in the video [27] we computed the average motion vector magnitude (MVM) over all the blocks and skin blocks. Additionally, we also computed the residual coding bitrate (RCB) that captures non-rigid, finer motion, such as hand gestures. Both these features capture different type of visible movement. The average of between the motion vectors and the residual coding bitrate was also calculated to capture visual activity. The features and their reference number are presented in Table 6.1.

Table 6.1: Visual Features

	Ref.	Label	Description
Visual Features	1	MVM_All	Motion Vector Magnitude computed over all blocks in the frame
	2	MVM_Skin	Motion Vector Magnitude computed over the skin blocks in the frame
	3	RCB_All	Residual Coding Bitrate computed on all the blocks
	4	RCB_Skin	Residual Coding Bitrate computed over the skin blocks in the frame
	5	MVMRCB_All_Av	Average between the Motion Vector Magnitude and Residual Coding Bitrate computed on all the blocks in the frame
	6	MVMRCB_Skin_Av	Average between the Motion Vector Magnitude and Residual Coding Bitrate computed on the skin blocks in the frame

6.2 Acoustic Features

The acoustic features for the human-human interaction dataset, were obtained from manually labeled data. Unlike in the human-computer interation condition, the automatic annotations of the speaker diarization system were not used to extract audio features. Due to the nature of our data (human-human interaction), all the particular characteristics of the speech signal, the quality of the audio signal, the results delivered by the speaker diarization system were not satisfactory. Having two human subjects, overlapping speech, proved to be a challenge for our system. Upon careful analysis of the automatic annotations on a subset of audio recordings that have been fed into the speaker diarization system, we realized we couldn't use these annotations for further feature processing. The large error recognition rates would mean that we have to manually correct the automatic speech segment annotations and this requires almost the same amount of time and effort. We therefore decided to pursue with the manual labeling of the data and extract the acoustic features from these annotations.

6.2.1 Manually Annotated Data

We manually labeled the speech segments of every subject and the experimenter using ANVIL [61]. The segments that contained speech, were labeled "Subject" or "Experimenter", depending on who was holding the floor; "Overlap" if the subject and experimenter were talking at the same time and "Silence", when none of the two were speaking. Figure 6.1 shows an instance of a subject's speech/silence segments annotation.

6.2.2 Acoustic Features from Manual Annotations

From the speaking activity annotations we derived the set of features related to the speaking behavior of the two participants involved in the interaction. The first four features in Table 6.2 are the features that were computed from the labeled data.

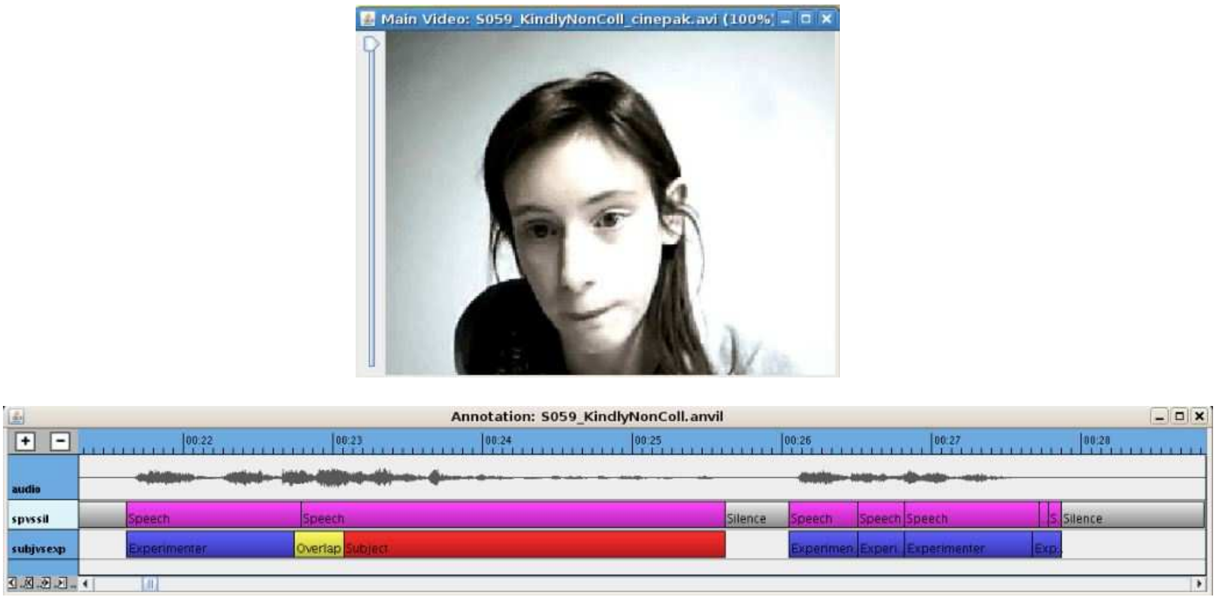


Figure 6.1: ANVIL snapshot of annotated speech and silence segments.

6.2.3 Acoustic Features Extracted Automatically

Keeping in line with the research methodology from the Self-Presentation and the Human-Computer Interaction chapters, acoustic features are derived automatically from properties of the pitch and intensity of the speech signal, using the audio processing software, Praat [17]. For this category of features, we only processed the subjects' speech segments. The pitch algorithm, based on the autocorellation method, and the processing parameters were kept the same: Praat's default values for time step and the pitch ceiling (600 Hz), lowering the pitch floor to 50 Hz, time step of 0.015 seconds and a window length of 0.06 seconds. Statistics such as median, average, standard deviation for pitch and intensity were computed.

Table 6.2 shows, apart from the features extracted from the labeled data, the automatically extracted features.

Table 6.2: Acoustic features

	Ref	Label	Description
Acoustic Features	7	DurSpeechExp	Duration of experimenter's speech
	8	DurSpeechSubj	Duration of subject's speech
	9	DurOverlap	Duration of the overlapping speech between the subject and the experimenter
	10	DurSilence	Duration of silence
	11	MeanPitch	Mean of subject's pitch
	12	MaxPitch	Maximum of subject's pitch
	13	MinPitch	Minimum of subject's pitch
	14	MedianPitch	Median of subject's pitch
	15	StDevPitch	Standard deviation of the subject's pitch
	16	MeanInt	Mean of the subject's speech intensity
	17	MaxInt	Maximum of the subject's speech intensity
	18	MinInt	Minimum of the subject's speech intensity
	19	StDevInt	Standard deviation of the subject's speech intensity

6.2.4 Additional Features

From the speech annotations, we derived two additional features that capture the dynamic of the dialogue: the total number of turns between the computer and the subject and the total number of turns between the computer and the subject, that last longer than 2 seconds. This way, possible events such as simple, short feedback utterances are eliminated. They are presented in Table 6.3.

Table 6.3: Additional Features

	Ref	Label	Description
Additional Features	20	SubjTurns	Number of speaking turns
	21	SubjLongTurns	Number of speaking turns longer than 2 seconds

6.3 Feature Analysis

The linear relationships between our features and the Big Five personality traits were analyzed by means of a number of backward linear regression analyses, one per each trait and per each collaboration level (CL), with all features as predictors.

Table 6.4 reports for each of the final models the retained predictors and the portion of dependent variable variance explained. Some of the linear models account only for a small portion of variance, while others account for a higher variance.

Table 6.7 (end of this Chapter, page 87) reports the partial correlations between the retained predictors and the various personality traits, in each of the 4 collaboration levels. Overall, compared to the Human-Computer Interaction case, the observed correlations between the traits and the features are sparse. In the first part of the interaction, when the experimenter is fully collaborative, people who produce more often longer speaking turns score higher on the Conscientiousness trait. Longer speaking durations are correlated with people higher on the Creativity scale. Keeping a constant voice tone and taking the floor often is positively correlated with Emotional Stability. People who move more score higher on Extraversion.

What is worth reporting from the two intermediate collaborative levels of interaction, A relatively constant pitch, an average and non-changing intensity and moving a lot is correlated with a higher score for Creativ-

Table 6.4: Retained predictors and portion of dependent variable variance explained.

Condition	Trait	Retained Predictors	R^2
CL1	Agreableness	3	.034
	Conscientiousness	20, 21	.016
	Creativity	8	.067
	Emot. Stability	7, 11, 14, 20, 21	.167
	Extraversion	3	.136
CL2	Agreableness	8, 9, 12, 15, 17, 2, 3, 6	.372
	Conscientiousness	7, 8, 13	.186
	Creativity	11, 15, 16, 17, 19, 1, 2, 3, 20, 6	.418
	Emot. Stability	13, 16, 17, 19, 6	.251
	Extraversion	8, 9, 10, 15, 17, 2, 3, 6	.385
CL3	Agreableness	15, 16, 17, 18, 1, 3, 6	.196
	Conscientiousness	12, 16, 17	.112
	Creativity	8, 9, 13, 14, 15, 18, 2	.386
	Emot. Stability	17, 19, 2, 6	.020
	Extraversion	10, 21	.170
CL4	Agreableness	12	.016
	Conscientiousness	1	.027
	Creativity	8, 15, 16, 1, 3, 6	.389
	Emot. Stability	8, 14, 15, 16, 17, 18, 1, 2, 6	.248
	Extraversion	10, 11, 14, 15, 20	.289

ity. People who speak with a low pitch but loud score higher on Emotional Stability. Extraversion is correlated with a high value for intensity and motion. A low intensity is positively correlated with Agreeableness. People who are more creative talk more, with a constant pitch. On Extraversion, those who score higher move more. In the fully non-collaborative interaction, a high value for pitch is correlated with Agreeableness. Longer speaking durations, in a constant voice tone, and display more movement, are found in those who are more creative. In this scenario, Extraversion is characterized by an average value for pitch and a high number of speaking turns.

6.4 Classification Experiments

6.4.1 Feature Selection

Before training, we apply a feature selection algorithm to nd the most informative features to avoid overtting. In our experiments, we used the Weka implementation of the Classifier Subset Evaluator that evaluates subsets of features using a classifier (in our case, Support Vector Machine with linear kernel) to estimate the 'merit' of a set of attributes [62]. The evaluation was performed on the datasets generated by an inner leave-one-subject-out cross validation. This method was used jointly with a Best First search strategy, that searches the space of subsets of features by greedy hill-climbing augmented with a backtracking facility. We used the forward direction of the search that begins at the empty set of features. In the Table 6.5, we report the features selected for the different traits using the linear SVM subset evaluator. These features were used for running the classification tasks.

6.4.2 Automatic Classification

For our classification experiments, all personality traits' scores were quantized (Low/High) along their median values (Agreeableness = 46.05, Conscientiousness = 49.57, Creativity = 47.48, Emotional Stability = 53.25, and Extraversion = 49.42). Classification was performed by means of Support Vector Machines (SVMs) with linear kernel. In particular, the bound-constrained SVM classification algorithm was used for the SVM classifier. The cost parameter C was estimated through an inner leave-one-subject-out cross validation on the training set of the first fold and were then kept fixed for the outer-cross validation.

We designed 5 binary classification tasks, one per personality trait. Each binary classification task was performed for each interaction modality, CL1

Table 6.5: Features selected using SVM - Classifier Subset Evaluator. Features' numbers refer to Table 6.1, Table 6.2 and Table 6.3

Cond.	Personality Trait				
	Extraversion	Agreableness	Conscientiousness	Emotional Stability	Creativity
CL1	6, 7, 8, 18	1, 2, 3, 9, 14, 15, 21	2, 4, 6, 7, 12, 16	1, 4, 6, 14	3, 5, 10
CL2	11, 17	7, 12	13, 14, 21	5, 10, 17	1, 3, 4, 8, 9, 13, 14, 16
CL3	1, 9, 13, 16	5, 7, 9, 10, 12, 13, 17	2, 4, 9, 17	1, 2, 4, 5, 7, 9, 18, 19	1, 4, 9, 10, 15, 16, 19
CL4	3, 9, 10, 11, 12	1, 10, 15, 17, 19	8, 9, 14, 18, 19	1, 3, 6, 17	1, 11, 18

- CL4. The total number of experimental conditions was 20. For each classifier, for each trait, and for each collaboration level of the interacting machine, only the features retained in the feature selection step (see Section 6.4.1) were used. The leave-one-out cross-validation strategy was employed. Hence, 31 models for each personality trait were trained on 32-subject subsets, evaluating them against the remaining ones and finally averaging the results. For comparison, we ran the same experiments described above by using SVMs with Radial Basis Function (RBF) kernels and Naïve Bayes. Since the classification experiments using SVMs with RBF kernels and Naïve Bayes gave results similar to those obtained with SVM with linear kernel, we only report the latter.

6.5 Classification Results

Table 6.6 summarizes the accuracy values for each trait and for each level of collaboration (CL), as described in Section 3.3, performed by the second participant. The accuracy values highlighted in bold are statistically significant, according to binomial tests that compared the observed accuracy to that of the baseline classifier that exploits the observed frequencies of the two classes. The significance was set at p$\leq$.01 and the resulting threshold values for accuracy was 0.70 for all the traits.

Table 6.6: Accuracy (%) obtained using SVM with linear kernel
.

	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Creativity
CL1	54.84	48.38	54.84	**77.42**	67.74
CL2	67.74	54.84	**74.19**	64.52	**74.19**
CL3	58.06	64.52	64.52	29.03	58.06
CL4	**71**	58.06	45.16	64.52	**71**

6.6 Discussion

Creativity seems to be the only trait yielding, over three out of four colaboration levels, results significantly better than the baseline. This trait is essential to our ability to adapt to the surrounding environment, handle unusual situations and it somehow influences how we perceive and interact with others in society. Studies about creativity focused on tests administered to students, asking them to draw a picture [106] or provide an alternative end for a known story or an event in history in order to assess their creativity level. Numerous research investigations have supported the importance of certain personality attributes, such as willingness to overcome obstacles for creative functioning [12], [42]. Hennessey discussed the intrinsic, task-focused motivation as being essential to creativity [52]. It follows, that the intrinsic motivation to successfully guide the experimenter through the map, reaching the end point is a possible explanation for why we observe the emergence of creativity in this setting.

A possible way of how the mechanism works, could be the following: the four interaction modalities (CL1 - CL4) somehow obliges the participant to look for alternative explanations when the experimenter directly or indirectly indicated that they have not reached the target or that they got lost. Since these events happen quite often during the entire interaction, explaining the same thing many times, but in a different manner can prove to be a challenge.

Apparently, trying to accomplish this task, in a fully collaborative and non-collaborative scenario, activates dispositions observable from vocal behavior. This is not unexpected, since there is no visual feedback from the experimenter. The behavioral signs that proved effective are: (i) pitch in three different conditions (CL1, CL2 and CL3) given by the particpant's role; (ii) duration of the experimenter's speech segments in the two intermediate levels (CL2, CL3) and the fully non-collaborative one (CL4); (iii) the duration of the overlapped speech segments in the one intermediate level (CL2) and the fully collaborative one (CL1);(iv) intensity (mean and standard deviation of the subject's voice intensity) in all the different conditions; and (v) body motion activity while the experimenter was interacting in an intermediate collaborative way (CL4).

A good accuracy was also obtained for Extraversion. The main descriptors of this trait revolve around the term "social". Extraverts tend to engage in social activities, they attract and enjoy social attention. A scenario like ours provides the a possible consequence of this view can be the explanation to why Extraversion-related disposition are activated to a greater extent. In particular, the behavioral signs associated with the accuracy's values that are statistically significant were: (i) intensity; (ii) the subject's and experimenter's overlapping speaking duration; and (iii) body motion activity.

Conscientiousness and Emotional Stability obtained a result slightly above the threshold only in the condition CL2, respectively CL1. In particular, for Agreeableness we could hypothesize that there is no coincidence that this trait is better recognized in the CL2 setting. The key might lie in the kind part of the experimenter while CL3 and CL4 might have no bearing: it activates a pleasing attitude in the subject.

Ultimately, we haven't identified any suitable setting/collaboration level to observe the emergence of behavioral descriptors associated with the Agreeableness trait.

Table 6.7: Partial correlations between the retained predictors and the personality traits; see Table 6.1, Table 6.2 and Table 6.3 for the features' reference numbers.

CL1	Features	13	-	-	-	-	-	-	-	-	-
	Agre.	-.257	-	-	-	-	-	-	-	-	-
	Features	20	21	-	-	-	-	-	-	-	-
	Consc.	-.295	.302	-	-	-	-	-	-	-	-
	Features	8	-	-	-	-	-	-	-	-	-
	Crea.	.314		-	-	-	-	-	-	-	-
	Features	7	11	14	20	21	-	-	-	-	-
	Em. Sta.	-.469	.359	-.353	.311	-.337	-	-	-	-	-
	Features	3	-	-	-	-	-	-	-	-	-
	Extr	.406	-	-	-	-	-	-	-	-	-
CL2	Features	8	9	12	15	17	2	3	6	-	-
	Agre.	.037	.011	.010	.003	.009	.044	.006	.037	-	-
	Features	7	8	13	-	-	-	-	-	-	-
	Consc.	.091	.031	.063	-	-	-	-	-	-	-
	Features	11	15	16	17	19	1	2	3	20	6
	Crea.	-.345	.449	.315	-.282	.335	-.317	.268	.464	-.291	-.281
	Features	13	16	17	19	6	-	-	-	-	-
	Em. Sta.	.316	-.339	.502	-.411	-.294	-	-	-	-	-
	Features	8	9	10	15	17	2	3	6	-	-
	Extr.	-.391	-.371	-.384	.294	.389	.439	.598	-.465	-	-
CL3	Features	15	16	17	18	1	3	6	-	-	-
	Agreeableness	-.309	-.442	.329	.499	-.284	.411	-.421	-	-	-
	Features	12	16	17	4	6	-	-	-	-	-
	Consc.	-.326	-.370	.388	-	-	-	-	-	-	-
	Features	8	9	13	14	15	18	2	-	-	-
	Crea.	.500	-.317	.334	-.333	.499	.476	-.448	-	-	-
	Features	17	19	2	6	2	3	4	-	-	-
	Em. Sta.	.323	-.332	.277	-.286	-	-	-	-	-	-
	Features	10	21	-	-	-	-	-	-	-	-
	Extr.	-.406	.372	-	-	-	-	-	-	-	-
CL4	Features	12	-	-	-	-	-	-	-	-	-
	Agre.	.232	-	-	-	-	-	-	-	-	-
	Features	1	-	-	-	-	-	-	-	-	-
	Consc.	.243	-	-	-	-	-	-	-	-	-
	Features	8	15	16	1	3	6	-	-	-	-
	Crea.	.255	.374	.505	-.454	.283	-.339	-	-	-	-
	Features	8	14	15	16	17	18	1	2	6	-
	Em. Sta.	.455	.295	.294	-.424	.389	.402	-.392	.346	-.411	-
	Features	10	11	14	15	20	-	-	-	-	-
	Extr.	-.458	.413	-.394	-.379	.404	-	-	-	-	-

7. Conclusion & Future work

The aim of this work is to contribute to the advancement of the state of the art for the automatic analysis of personality given three different scenarios. In particular, we investigate the feasibility of detecting the Big Five in a) short videos of self-presentation; b) in the context of Human-Computer Interaction; and c) in the context of Human-Human Interaction. We adopted a thin-slice perspective and extracted a set of acoustic and visual features using supervised machine learning algorithms to investigate the contribution of the large set of acoustic and visual non-verbal cues, for the classfication task.
An extensive literature review "shows", at the same time, the missing pieces in the emerging field of personality assessment and also the advantages of applications that actively use dedicated algorithms.

In Chapter 4, we investigated the automatic recognition of the Big Five personality dimensions from self-presentation videos. After recording 93 videos, 89 of which were used, in which participants were asked to briefly introduce themselves, we extracted a number of visual and acoustic features that have been shown to contain salient information with respect to the participant's personality. The audio cues were based on automatic computation of the pitch and intensity of the voice signal. The visual cues were computed from manual annotations of the eye-gaze, hand movement, head orientation, body posture and mouth fidgeting.
The main findings of our work in this setting are the following:

a) Conscientiousness and Emotional Stability are the easiest traits to automatically detect during self-presentation. The reason could be that the first trait is connected to engagement within task-related behavior and that the second is connected to the emotional reactions (e.g. distress) it elicits.

b) Our task does not seem to activate the full range of dispositions of Agreeableness and Extraversion. For Extraversion, the reasons can be that introducing oneself in front of a computer screen does not provide enough social audience to let the social attention dispositions of extraverts fully activate. As to Agreeableness, we have invoked the masking effect of the necessity of pleasing the experimenter, implicit in the nature of the situation.

On the practical side, the results we obtained in this setting are an important first step towards automatic systems assisting either interviewers or interviewees in improving their performance in job interviews. On a more theoretical side, they emphasize the influence of the situation for a full unfolding of the behavioral dispositions tied to personality traits.

Understandably, more work is needed to fully explore the automatic analysis of personality traits in self-introduction, e.g., by considering even larger sets of non-verbal features, as well as verbal ones (e.g., lexical choice, presence of emotion-related words; topic dynamics, etc.); using larger samples and/or exploiting regression or ordinal techniques. Finally, another possibility worth considering is the possibility of extending the work to interviewer/interviewee interactions by collecting new data to also model this scenario, where the system can work with the mediation of an interviewer.

The second part of the work, presented in Chapter 5, deals with a more interactive setting: Human-Computer Interaction. In this section, we investigated the emergence of the five personality traits, in HCI setting, with the interaction taking place under four different collaboration levels. These collaboration levels were used in order to elicit the manifestation of behaviors typical to trait, thus making a trait more "observable" and able to be evidenced by visual and acoustic analysis. Each interaction was subjected to the same chronological order: starting with a fully collaborative level (CL1), going through two intermediate (CL2 and CL3) and ending with a fully non-collaborative one (CL4). We analyzed the contribution of a large set of different acoustic and visual non-verbal cues. The main findings of our work in this setting are the following:

a) Emotional Stability and Extraversion are the easiest traits to automatically detect under the different collaborative settings (CL1, CL2, CL3, CL4 for Emotional Stability and CL2, CL3, and CL4 for Extraversion). For the Extraversion, the reason could be that the setting contains sufficient social "ingredients", thus allowing us to better observe and capture the characteristics of the trait, while for the latter, which is connected to emotional reactions (e.g. distress), the interaction and the various behaviors of the machine, elicits its manifestation.

b) Our task does not seem to activate the full range of dispositions for Creativity. We have invoked the masking effect of the necessity of giving precise indications to the experimenter, implicit in the nature of the situation.

Interestingly, Agreeableness and Conscientiousness, perform better than the baseline, only under a moderately non-collaborative setting. Since the main characteristics of the Agreeableness trait are related to pleasing the other interacting person, and receiving attention and it's not a trait that exhibits antipodal behaviors, a possible explanation could be that opposite, extreme settings (fully collaborative or fully non-collaborative) do not foster conditions for a clear observation of the trait. For Conscientiousness, it is possible that only CL2-like settings are appropriate for further studies on this trait.

Although not excellent, part of our results are encouraging and represent an important initial step towards automatic systems capable of recognizing personality under different circumstances and in different scenarios (e.g. assistive robotics, human-interaction collaborative systems). On a more theoretical side, they emphasize and demonstrate up to a certain point, the influence of the situation for a full unfolding of the behavioral dispositions tied to personality traits.
The original hypotheses, that differences in the interaction context are associated with differences in the way personality traits manifest, and that these can be used to better predict the traits, could not be proved to meet our original expectations. We believe that the limitation lies in sample size of our data. This could have influenced our results by limiting the scientific observations which could have been possible to make with the behavioral conditions we employed for this paper.

More work is needed to fully explore the automatic analysis of personality traits under the considered conditions. A continuing strategy would be have a richer dataset or to consider even larger sets of non-verbal feature, as well as verbal ones (e.g., lexical choice, presence of emotion-related words; topic dynamics, etc.); and/or exploiting regression or ordinal techniques in the analysis of the data.

Our last scenario, presented in Chapter 6, deals with a complementary social setting: Human-Human Interaction. The work on this scenario involves the same research methodology used in the Human-Computer Interaction scenario. The steps regarding multimodal data processing, feature extraction and running the classification algorithms were the same as in Chapter 4 with a minor exception: the dialogue between the subject and the experimenter was labeled manually and not automatically. Results have In this scenario, we were able to identify different personality traits, as being best observable in this setting than in the Human-Computer Interaction. Under close inspection, it seems that this scenario

The main findings of our work in the Human-Human Interaction setting are:

a) Creativity is the easiest trait to automatically detect under the different collaborative settings (CL1, CL2, and CL4). Given the nature of the interaction and the goal-oriented task, low-level descriptors, characteristic of Creativity can surface and be captured by different fetaures.

b) Extraversion is the second best observable trait. Unsurprisingly Extraversion unfolds best in a social setting and has a great impact on our social behaviour.

c) This scenario in combination with the task does not seem to set off the full range of dispositions for Agreeableness. Based on the small data sample, it is also possible to have insufficient data to capture low-level behaviors typical for this trait.

Interestingly, Agreeableness and Conscientiousness, perform better than the baseline, only under a moderately non-collaborative setting. Since the main characteristics of the Agreeableness trait are related to pleasing the other interacting person, and receiving attention and it's not a trait that exhibits antipodal behaviors, a possible explanation could be that opposite, extreme settings (fully collaborative or fully non-collaborative) do not foster conditions for a clear observation of the trait. For Conscientiousness, it is possible that only CL2-like settings are appropriate for further studies on this trait.

Although not excellent, part of our results are encouraging and represent an important initial step towards automatic systems capable of recognizing personality under different circumstances and in different scenarios (e.g. assistive robotics, human-interaction collaborative systems). On a more theoretical side, they emphasize and demonstrate up to a certain point, the influence of the situation for a full unfolding of the behavioral dispositions tied to personality traits.
The original hypotheses, that differences in the interaction context are associated with differences in the way personality traits manifest, and that these can be used to better predict the traits, could not be proved to meet our original expectations. We believe that the limitation lies in sample size of our data. This could have influenced our results by limiting the scientific observations which could have been possible to make with the behavioral conditions we employed for this paper.

More work is needed to fully explore the automatic analysis of personality traits under the considered conditions. A continuing strategy would be have a richer dataset or to consider even larger sets of non-verbal feature, as well as verbal ones (e.g., lexical choice, presence of emotion-related words; topic dynamics, etc.); and/or exploiting regression or ordinal techniques in the analysis of the data.
We can say that this scenario captures the variability in human behavior in a different manner than the Human-Computer Interaction scenario does. Our work also has the advantage that by aggregating the data and findings from the Human-Computer and Human-Human Interaction, we can have a solid, consensus view to describe the probability of how likely it is to recognize personality in any given setting. The occurring behavior

better captures the variability in human behavior,thus the personality in a different manner.

The overall conclusion is that our studies have shown the feasibility of automatically assessing personality traits based on thin slices of behaviour and have indicates which features and which scenario or context is more appropriate for personality emergence and ultimately assessment.

Bibliography

[1] Nonlinear support vector machines. http://www.bindichen.co.uk/post/AI/Nonlinear-Support-Vector-Mach Accessed April 4, 2010.

[2] G. Allport. *Pattern and growth in personality.* Holt, Rinehart and Winston, 1961.

[3] Nalini Ambady, Mark Hallahan, and Brett Conner. Accuracy of judgments of sexual orientation from thin slices of behavior. *Journal of personality and social psychology*, 77(3):538, 1999.

[4] Nalini Ambady and Robert Rosenthal. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological bulletin*, 111(2):256, 1992.

[5] Nalini Ambady and Robert Rosenthal. Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of personality and social psychology*, 64(3):431, 1993.

[6] A. Anderson, M. Bader, E. Bard, E. Boyle, G.M. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H.S. Thompson, and R. Weinert. The hcrc map task corpus. *Annual Review of Psychology*, 34:351–366, 1991.

[7] Elisabeth André, Martin Klesen, Patrick Gebhard, Steve Allen, and Thomas Rist. Integrating models of personality and emotions into lifelike characters. In *Affective interactions*, pages 150–165. Springer, 2000.

[8] Shlomo Argamon, Sushant Dhawle, Moshe Koppel, and James Pennebaker. Lexical predictors of personality type. 2005.

[9] Charles D Aronovitch. The voice of personality: Stereotyped judgments and their relation to voice quality and sex of speaker. *The Journal of Social Psychology*, 99(2):207–220, 1976.

[10] M. C. Ashton and S. V. Lee, K. & Paunonen. What is the central feature of extraversion?: Social attention versus reward sensitivity. *Journal of Personality and Social Psychology*, 83(1):245–251, 2002.

[11] Michael C Ashton and Kibeom Lee. A theoretical basis for the major dimensions of personality. *European Journal of Personality*, 15(5):327–353, 2001.

[12] Frank Barron. Creative person and creative process. 1969.

[13] Ligia Batrinca, Giota Stratou, Ari Shapiro, Louis-Philippe Morency, and Stefan Scherer. Cicero-towards a multimodal virtual audience platform for public speaking training. In *Intelligent Virtual Agents*, pages 116–128. Springer Berlin Heidelberg, 2013.

[14] Ligia Maria Batrinca, Nadia Mana, Bruno Lepri, Fabio Pianesi, and Nicu Sebe. Please, tell me about yourself: automatic personality assessment using short self-presentations. In *Proceedings of the 13th international conference on multimodal interfaces*, pages 255–262. ACM, 2011.

[15] Timothy W Bickmore and Rosalind W Picard. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(2):293–327, 2005.

[16] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[17] Paul Boersma and David Weenink. Praat software. *Amsterdam: University*, 2006.

[18] Peter Borkenau and Anette Liebler. Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, 62(4):645, 1992.

[19] Peter Borkenau, Nadine Mauer, Rainer Riemann, Frank M Spinath, and Alois Angleitner. Thin slices of behavior as cues of personality and intelligence. *Journal of personality and social psychology*, 86(4):599, 2004.

[20] Jr. Bouchard, Thomas J. and John C. Loehlin. Genes, evolution, and personality. *Behavior Genetics*, 31(3):243–273, 2001.

[21] John Brebner. Personality theory and movement. In *Individual differences in movement*, pages 27–41. Springer, 1985.

[22] Christopher JC Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.

[23] Avshalom Caspi, Brent W Roberts, and Rebecca L Shiner. Personality development: Stability and change. *Annual Review of Psychology*, 56:453–484, 2005.

[24] Justine Cassell and Timothy Bickmore. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. user modeling and adaptive interfaces. pages 89–132, 2003.

[25] Federica Cavicchio and Massimo Poesio. Annotation of cooperation and emotions in map task dialogues. *Programme of the Workshop on Multimodal Corpora*, page 1, 2008.

[26] Gokul Chittaranjan, Jan Blom, and Daniel Gatica-Perez. Mining large-scale smartphone data for personality studies. *Personal and Ubiquitous Computing*, 17(3):433–450, 2013.

[27] Miguel T Coimbra and Mike Davies. Approximating optical flow within the mpeg-2 compressed domain. *Circuits and Systems for Video Technology, IEEE Transactions on*, 15(1):103–107, 2005.

[28] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[29] P. T. Costa and R. R. McCrae. *The NEO Personality Inventory manual*, 1985.

[30] Paul Costa Jr, Antonio Terracciano, and Robert R McCrae. Gender differences in personality traits across cultures: robust and surprising findings. *Journal of personality and social psychology*, 81(2):322, 2001.

[31] Mark Costanzo and Dane Archer. Interperting the expressive behavior of others: The interpersonal perception task. *Journal of Nonverbal Behavior*, 13(4):225–245, 1989.

[32] L. I. Cripe. *Personality assessment of brain-impaired patients.* Lawrence Erlbaum Associates, 1996.

[33] Jared R Curhan and Alex Pentland. Thin slices of negotiation: Predicting outcomes from conversational dynamics within the first 5 minutes. *Journal of Applied Psychology*, 92(3):802, 2007.

[34] Rodrigo de Oliveira, Alexandros Karatzoglou, Pedro Concejero Cerezo, Ana Armenta Lopez de Vicuña, and Nuria Oliver. Towards a psychographic user model from mobile phone usage. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 2191–2196. ACM, 2011.

[35] Fiorella De Rosis and Cristiano Castelfranchi. How can personality factors contribute to make agents more believable. In *Proceedings of the I3 Spring Days Workshop onBehaviour Planning for Lifelike Characters and Avatars*. Citeseer, 1999.

[36] Timothy DeGroot and Janaki Gooty. Can nonverbal cues be used to make meaningful personality attributions in employment interviews? *Journal of Business and Psychology*, 24(2):179–192, 2009.

[37] Jean-Marc Dewaele and Adrian Furnham. Personality and speech production: a pilot study of second language learners. *Personality and Individual Differences*, 28(2):355–365, 2000.

[38] Alexandra Ehrenberg, Suzanna Juckes, Katherine M White, and Shari P Walsh. Personality and self-esteem as predictors of young people's technology use. *CyberPsychology & Behavior*, 11(6):739–741, 2008.

[39] Nicole Ellison, Rebecca Heino, and Jennifer Gibbs. Managing impressions online: Self-presentation processes in the online dating environment. *Journal of Computer-Mediated Communication*, 11(2):415–441, 2006.

[40] David Feil-Seifer and Maja J Mataric. Defining socially assistive robotics. In *Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on*, pages 465–468. IEEE, 2005.

[41] Alan Feingold. Gender differences in personality: a meta-analysis. *Psychological bulletin*, 116(3):429, 1994.

[42] Gregory J Feist. The function of personality in creativity. *The Cambridge handbook of creativity*, pages 113–130, 2010.

[43] David C Funder and Carl D Sneed. Behavioral manifestations of personality: an ecological approach to judgmental accuracy. *Journal of personality and social psychology*, 64(3):479, 1993.

[44] Adrian Furnham. Language and personality. 1990.

[45] Adrian Furnham, Chris J Jackson, and Tony Miller. Personality, learning style and work performance. *Personality and Individual Differences*, 27(6):1113 – 1122, 1999.

[46] Robert Gifford, Cheuk Fan Ng, and Margaret Wilkinson. Nonverbal cues in the employment interview: Links between applicant qualities and interviewer judgments. *Journal of Applied Psychology*, 70(4):729, 1985.

[47] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1-3):389–422, 2002.

[48] Jeffrey A Hall, Namkee Park, Hayeon Song, and Michael J Cody. Strategic misrepresentation in online dating: The effects of gender, self-monitoring, and personality traits. *Journal of Social and Personal Relationships*, 27(1):117–135, 2010.

[49] Jeffrey T Hancock, Lauren E Curry, Saurabh Goorha, and Michael Woodworth. On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes*, 45(1):1–23, 2007.

[50] Fumio Hara. Artificial emotion of face robot through learning in communicative interactions with human. In *Robot and Human Interactive Communication, 2004. ROMAN 2004. 13th IEEE International Workshop on*, pages 7–15. IEEE, 2004.

[51] Alan D. Mead Heather E.P. Cattell. *The Sixteen Personality Factor Questionnaire (16PF)*, pages 135–160. SAGE Publications Ltd, 0 edition, 2008.

[52] Beth A Hennessey. The creativity-motivation connection. *The Cambridge handbook of creativity*, pages 342–365, 2010.

[53] Deniz S. Hogan, Joyce & Ones. *Conscientiousness and integrity at work.*, pages 849–870. San Diego, CA, US: Academic Press, 0 edition, 1997.

[54] Robert Hogan, Gordon J Curphy, and Joyce Hogan. What we know about leadership: Effectiveness and personality. *American psychologist*, 49(6):493–504, 1994.

[55] Allen I Huffcutt, James M Conway, Philip L Roth, and Nancy J Stone. Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology*, 86(5):897, 2001.

[56] F Iida, M Tabata, and F Hara. Generating personality character in a face robot through interaction with human. In *7th IEEE In-*

ternational Workshop on Robot and Human Communication, pages 481–486, 1998.

[57] Paul Ingram and Michael W Morris. Do people mix at mixers? structure, homophily, and the life of the party. *Administrative Science Quarterly*, 52(4):558–585, 2007.

[58] Dinesh Babu Jayagopi, Hayley Hung, Chuohao Yeo, and Daniel Gatica-Perez. Modeling dominance in group conversations using nonverbal activity cues. *Audio, Speech, and Language Processing, IEEE Transactions on*, 17(3):501–513, 2009.

[59] Oliver P John, Laura P Naumann, and Christopher J Soto. Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research*, 3:114–158, 2008.

[60] Oliver P John and Sanjay Srivastava. *The Big Five trait taxonomy: History, measurement, and theoretical perspectives*, volume 2, pages 102–138. 1999.

[61] Michael Kipp. Anvil 3.5. 2002.

[62] Ron Kohavi and George H John. Wrappers for feature subset selection. *Artificial intelligence*, 97(1):273–324, 1997.

[63] Meera Komarraju, Steven J. Karau, and Ronald R. Schmeck. Role of the big five personality traits in predicting college students' academic motivation and achievement. *Learning and Individual Differences*, 19(1):47 – 52, 2009.

[64] Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805, 2013.

[65] Wilburn Lane and Chris Manner. The impact of personality traits on smartphone ownership and use. *International Journal of Business and Social Science*, 2(17):22–28, 2011.

[66] Cyril Laurier, Jens Grivolla, and Perfecto Herrera. Multimodal music mood classification using audio and lyrics. In *Machine Learning and Applications, 2008. ICMLA'08. Seventh International Conference on*, pages 688–693. IEEE, 2008.

[67] Mark R Leary and Ashley Batts Allen. Personality and persona: Personality processes in self-presentation. *Journal of personality*, 79(6):1191–1218, 2011.

[68] Kwan Min Lee, Wei Peng, Seung-A Jin, and Chang Yan. Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in humanrobot interaction. *Journal of Communication*, 56(4):754–772, 2006.

[69] Bruno Lepri, Nadia Mana, Alessandro Cappelletti, Fabio Pianesi, and Massimo Zancanaro. Modeling the personality of participants during group interactions. In *User Modeling, Adaptation, and Personalization*, pages 114–125. Springer, 2009.

[70] Bruno Lepri, Jacopo Staiano, Giulio Rigato, Kyriaki Kalimeri, Ailbhe Finnerty, Fabio Pianesi, Nicu Sebe, and Alex Pentland. The sociometric badges corpus: A multilevel behavioral dataset for social behavior in complex organizations. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pages 623–628. IEEE, 2012.

[71] Bruno Lepri, Ramanathan Subramanian, Kyriaki Kalimeri, Jacopo Staiano, Fabio Pianesi, and Nicu Sebe. Employing social gaze and

speaking activity for automatic determination of the extraversion trait. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, page 7. ACM, 2010.

[72] Jackson Liscombe, Jennifer Venditti, and Julia Hirschberg. Classifying subject ratings of emotional speech using acoustic features. Proceedings of Eurospeech 2003, 2003.

[73] Bing Liu. Sentiment analysis and subjectivity. *Handbook of natural language processing*, 2:568, 2010.

[74] Richard E Lucas, Kimdy Le, and Portia S Dyrenforth. Explaining the extraversion/positive affect relation: Sociability cannot account for extraverts' greater happiness. *Journal of personality*, 76(3):385–414, 2008.

[75] François Mairesse, Marilyn A. Walker, Matthias R. Mehl, and Roger K. Moore. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research, Vol*, 30:457–501, 2007.

[76] J. D. Mayer. *A classification of DSM-IV-TR mental disorders according to their relation to the personality system. Comprehensive handbook of personality and psychopathology (CHOPP) Vol. 1: Personality and everyday functioning.* John Wiley & Sons, 2005.

[77] Emma Medford and Sarah P. McGeown. The influence of personality characteristics on children's intrinsic reading motivation. *Learning and Individual Differences*, 22(6):786 – 791, 2012.

[78] Roland Mergl, Michael Vogel, Anuschka Prässl, Birgit Graf, Max Karner, Paraskevi Mavrogiorgou, Ulrich Hegerl, and Georg Juckel.

Facial expressions and personality: A kinematical investigation during an emotion induction experiment. *Neuropsychobiology*, 54(2):114–119, 2007.

[79] Gilad Mishne. Experiments with mood classification in blog posts. In *Proceedings of ACM SIGIR 2005 Workshop on Stylistic Analysis of Text for Information Access*, volume 19, 2005.

[80] Gelareh Mohammadi and Alessandro Vinciarelli. Automatic personality perception: Prediction of trait attribution based on prosodic features. 2012.

[81] Gelareh Mohammadi, Alessandro Vinciarelli, and Marcello Mortillaro. The voice of personality: mapping nonverbal vocal behavior into trait attributions. In *Proceedings of the 2nd international workshop on Social signal processing*, pages 17–20. ACM, 2010.

[82] Greg Murray, David Rawlings, Nicholas B Allen, and John Trinder. Neo five-factor inventory scores: Psychometric properties in a community sample. *Measurement and Evaluation in Counseling and Development*, 2003.

[83] Andreas Mhlberger, Martin J Herrmann, Georg Wiedemann, Heiner Ellgring, and Paul Pauli. Repeated exposure of flight phobics to flights in virtual reality. *Behaviour Research and Therapy*, 39(9):1033 – 1050, 2001.

[84] Hiroshi Nakajima, Ryota Yamada, Scott Brave, Yasunori Morishima, Clifford Nass, and Shigeyasu Kawaji. The functionality of human-machine collaboration systems-mind model and social behavior. In *Systems, Man and Cybernetics, 2003. IEEE International Conference on*, volume 3, pages 2381–2387. IEEE, 2003.

[85] Clifford Nass and Kwan Min Lee. Does computer-synthesized speech manifest personality? experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied*, 7(3):171, 2001.

[86] Clifford Nass, Youngme Moon, BJ Fogg, Byron Reeves, and Chris Dryer. Can computer personalities be human personalities? In *Conference companion on Human factors in computing systems*, pages 228–229. ACM, 1995.

[87] Clifford Ivar Nass and Scott Brave. *Wired for speech: How voice activates and advances the human-computer relationship.* MIT press Cambridge, 2005.

[88] Jon Oberlander and Scott Nowson. Whose thumb is it anyway?: classifying author personality from weblog text. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 627–634. Association for Computational Linguistics, 2006.

[89] Daniel Olguín Olguín, Peter A Gloor, and Alex Pentland. Wearable sensors for pervasive healthcare management. In *Pervasive Computing Technologies for Healthcare, 2009. PervasiveHealth 2009. 3rd International Conference on*, pages 1–4. IEEE, 2009.

[90] Pierre-yves Oudeyer. Novel useful features and algorithms for the recognition of emotions in human speech. In *Speech Prosody 2002, International Conference*, 2002.

[91] Jamie Pearson, Jiang Hu, Holly P Branigan, Martin J Pickering, and Clifford I Nass. Adaptive language behavior in hci: how expectations and beliefs about a system affect users' word choice. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1177–1180. ACM, 2006.

[92] Marco Perugini and L Di Blas. The big five marker scales (bfms) and the italian ab5c taxonomy: Analyses from an emic-etic perspective. 2002.

[93] Fabio Pianesi, Nadia Mana, Alessandro Cappelletti, Bruno Lepri, and Massimo Zancanaro. Multimodal recognition of personality traits in social interactions. In *Proceedings of the 10th international conference on Multimodal interfaces*, pages 53–60. ACM, 2008.

[94] Susan M Raza and Bruce N Carpenter. A model of hiring decisions in real employment interviews. *Journal of Applied Psychology*, 72(4):596, 1987.

[95] Byron Reeves and Clifford Nass. *The media equation: how people treat computers, television, and new media like real people and places.* Cambridge University Press, New York, NY, USA, 1996.

[96] Rutger Rienks and Dirk Heylen. Dominance detection in meetings using easily obtainable features. In *Machine Learning for Multimodal Interaction*, pages 76–86. Springer, 2006.

[97] Ronald E Riggio and Howard S Friedman. Impression formation: The role of expressive behavior. *Journal of Personality and Social Psychology*, 50(2):421, 1986.

[98] Robert Rosenthal, Judith A Hall, M Robin DiMatteo, Peter L Rogers, and Dane Archer. *Sensitivity to nonverbal communication: The PONS test.* Johns Hopkins University Press Baltimore, 1979.

[99] Philip L Roth, Chad H Iddekinge, Allen I Huffcutt, Carl E Eidson, and Mark J Schmit. Personality saturation in structured interviews. *International Journal of Selection and Assessment*, 13(4):261–273, 2005.

[100] Barbara Olasov Rothbaum, Page Anderson, Elana Zimand, Larry Hodges, Delia Lang, and Jeff Wilson. Virtual reality exposure therapy and standard (in vivo) exposure therapy in the treatment of fear of flying. *Behavior Therapy*, 37(1):80–90, 2006.

[101] Klaus Rainer Scherer. *Personality markers in speech.* Cambridge University Press, 1979.

[102] David P Schmitt, Jüri Allik, Robert R McCrae, and Verónica Benet-Martínez. The geographic distribution of big five personality traits patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology*, 38(2):173–212, 2007.

[103] Gwendolyn Seidman. Self-presentation and belonging on facebook: How personality influences social media use and motivations. *Personality and Individual Differences*, 54(3):402 – 407, 2013.

[104] Michelle N Shiota, Dacher Keltner, and Oliver P John. Positive emotion dispositions differentially associated with big five personality and attachment style. *The Journal of Positive Psychology*, 1(2):61–71, 2006.

[105] Jon F Sigurdsson. Computer experience, attitudes toward computers and personality characteristics in psychology undergraduates. *Personality and Individual Differences*, 12(6):617–624, 1991.

[106] Robert J Sternberg and Todd I Lubart. *Defying the crowd: Cultivating creativity in a culture of conformity.* Free Press, 1995.

[107] Adriana Ţăpuş, Cristian Ţăpuş, and Maja J Matarić. Userrobot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Intelligent Service Robotics*, 1(2):169–183, 2008.

[108] Ian H Witten, Eibe Frank, and Mark A Hall. *Data Mining: Practical machine learning tools and techniques.* Morgan Kaufmann, 2011.

[109] Chuohao Yeo and Kannan Ramchandran. Compressed domain video processing of meetings for activity estimation in dominance classification and slide transition detection. *University of California, Berkeley, Tech. Rep. UCB/EECS-2008-79*, 2008.

[110] Steve Young, Gunnar Evermann, Dan Kershaw, Gareth Moore, Julian Odell, Dave Ollason, Valtcho Valtchev, and Phil Woodland. The htk book. *Cambridge University Engineering Department*, 3:175, 2002.

[111] Gloria Zen, Bruno Lepri, Elisa Ricci, and Oswald Lanz. Space speaks: towards socially and personality aware visual surveillance. In *Proceedings of the 1st ACM international workshop on Multimodal pervasive video analysis*, pages 37–42. ACM, 2010.

A. Appendix

The personality questionnaire administered to the participants is presented in Table A.1. Each personality trait has 10 items and each item is evaluated on a 7 - point Likert scale.

Computing the Raw Scores: The scores for the adjectives in the second column were reversed - a score of 7 became 1, 6 became 2 and so on. The scale was from 1 - strongly disagree to 7 - strongly agree). The scores were then averaged across trait for each person (10 adjectives).

Computing the Z - Scores: These scores were computed using this formula: $(Rawscore) - (Average)/(StandardDeviation)$. They were calculated for all five personality traits. Average and standard deviation were computed according to our population. Scores were dependent on the gender of the subject and their age. A different score was obtained, whether the participant was male and under(young) or over(adult) the age 25 or female and under(young) or over(adult) age 25.

Computing the Orthogonalized Scores: The Z-scores were converted to Orthogonalised scores. This was done by taking the Z-score for each of the personality traits and using a formula from [92]:

Extraversion=(1.051**Extr*)-(0.077**Agre*)+(0.069**Consc*)-(0.057**EmSt*)-(0.149**Crea*)

Agreeableness=-(0.075**Extr*)+(1.022**Agre*)-(0.042**Consc*)-(0.067**EmSt*)-(0.024**Crea*)

Conscientiousness=(0.071**Extr*)-(0.043**Agre*)+(1.062**Consc*)-(0.196**EmSt*)+(0.012**Crea*)

Em. Stability=-(0.059**Extr*)-(0.07**Agre*)-(0.197**Consc*)+(1.066**EmSt*)+(0.015**Crea*)

Creativity=(0.147**Extr*)-(0.024**Agre*)+(0.011**Consc*)+(0.015**EmSt*)+(1.034**Crea*)

The italicised traits represent the Z-Score for that particular personality trait, such that each Final Score for Extraversion, for example, consists of a combination of each personality trait.
The Orthogonalised scores were standardised with the following formula $(PersonalityTrait * 10) + 50$.

Table A.1: Personality Questionnaire

Extraversion	
Extraverted (Estroverso)	Reserved (Riservato)
Warmhearted (Espansivo)	Shy (Timido)
Open (Aperto)	Silent (Silenzioso)
Exuberant (Esuberante)	Introverted (Introverso)
Vivacious (Vivace)	Reserved (Chiuso)
Agreeableness	
Altruistic (Altruista)	Egoistic (Egoista)
Agreeable (Disponibile)	Revengeful (Vendicativo)
Generous (Generoso)	Cynical (Cinico)
Sympathetic (Comprensivo)	Egocentric (Egocentrico)
Hospitable (Ospitale)	Suspicious (Sospettoso)
Conscientiousness	
Precise (Preciso)	Untidy (Disordinato)
Orderly (Ordinato)	Inconstant (Incostante)
Diligent (Diligente)	Careless (Impreciso)
Methodical (Metodico)	Careless (Sbadato)
Conscientious (Coscienzioso)	Rash (Incosciente)
Emotional Stability	
Self-assured (Sicuro)	Nervous (Nervoso)
Serene (Sereno)	Anxious (Ansioso)
Calm (Calmo)	Emotional (Emotivo)
Impassive (Impassibile)	Susceptible (Suscettibile)
Jealous (Geloso)	Touchy (Permaloso)
Creativity	
Creative (Creativo)	Superficial (Superficiale)
Imaginative (Fantasioso)	Obtuse (Ottuso)
Original (Originale)	Ingenious (Ingegnioso)
Poetic (Poetico)	Intuitive (Intuitivo)
Intelligent (Intelligente)	Rebellious (Ribelle)