The Voice of Personality: Mapping Nonverbal Vocal Behavior into Trait Attributions

Gelareh Mohammadi
Idiap Research Institute
CP 592 - 1920 Martigny (CH)
École Polytechnique Fédérale
de Lausanne - EPFL
1015-Lausanne (CH)
gelareh.mohammadi@epfl.ch

Alessandro Vinciarelli
University of Glasgow
G12 8QQ Glasgow (UK)
Idiap Research Institute
CP 592 - 1920 Martigny (CH)
vincia@dcs.gla.ac.uk

Marcello Mortillaro
University of Geneva
Rue des Battoirs 7
1205 Geneva (CH)
marcello.mortillaro@unige.ch

ABSTRACT

This paper reports preliminary experiments on automatic attribution of personality traits based on nonverbal vocal behavioral cues. In particular, the work shows how prosodic features can be used to predict, with an accuracy up to 75% depending on the trait, the personality assessments performed by human judges on a collection of 640 speech samples. The assessments are based on a short version of the Big Five Inventory, one of the most widely used questionnaires for personality assessment. The judges did not understand the language spoken in the speech samples so that the influence of the verbal content is limited. To the best of our knowledge, this is the first work aimed at inferring automatically traits attributed by judges rather than traits self-reported by subjects.

Categories and Subject Descriptors: H.3.1 [Content Analysis and Indexing].**General Terms:** Experimentation. **Keywords:** Personality Assessment, Big Five Personality Mode, Social Signal Processing, Nonverbal Vocal Behavior

1. INTRODUCTION

Whenever we meet unknown persons, we make spontaneous inferences about a wide range of socially relevant characteristics including attitudes, intentions, values and beliefs [19]. This work considers one facets of this phenomenon, namely the spontaneous and immediate attribution of personality traits to unknown individuals. The process can be very fast (100 *ms* have been shown to be sufficient for attributing competence to face images [17]) and it does not necessarily identifies the actual traits of a person. However, attributed traits show how an individual is *perceived* and this is important because it is the way we perceive others that drives our attitudes and behaviors towards them.

More specifically, this paper investigates the inference of personality traits from nonverbal features of speech. Our study shows that the personality assessments made by human judges can be automatically inferred from prosodic features extracted directly from the speech signal. The personality assessments are made with a short version (10 items) of the Big Five Inventory (BFI) [13], one of the most common personality assessment tools. The results show that the traits attributed by human judges can be predicted with an accuracy ranging from 65% to 80%.

Following the Social Signal Processing framework [21], we decided to use speech prosodic features, as they do not take into account what people say, but rather how they say it. These features may have a special role in the case of spontaneous trait attribution because this phenomenon takes place in the first few seconds after an unknown individual is met, sometimes even before that the verbal content of a message is fully delivered and can influence the perception of a listener. To further highlight this effect, the assessments have been performed over clips of language unknown to the judges. In this way, participants do not understand what the speakers say and can base their judgements only on speaker's style of speech.

To the best of our knowledge, the only other approach aimed at predicting others personality assessments is in [8]. However, such work includes lexical features while this one is based on purely nonverbal behavioral cues. Other personality related works have concentrated on individual traits such as extroversion or locus of control [9][12], or they have measured the correlation between nonverbal cues and personality traits, without performing automatic inference or prediction [10].

In the human sciences community the attribution of personality traits has been investigated more extensively, but speech and voice have been partly neglected with respect to other sources of trait inference like faces [17] or descriptions of behavior [19]. To the best of our knowledge, after some pioneering investigations in the late seventies [15][16], the presence of proximal and distal cues of personality in speech has not been the subject of major research efforts.

Approaches like those presented in this work can be beneficial for several technological domains. In multimedia indexing and retrieval, the automatic attribution of personality traits can help to better understand how data consumers perceive the content of data portraying people. This will make multimedia retrieval systems more adapted to human needs [11]. In Human-Computer Interaction, findings about the way speech elicits personality perceptions can help to

SSPW'10, October 29, 2010, Firenze, Italy.

build more accepted speaking machines like, e.g., GPS and automatic dialogue systems [9]. In speech synthesis, the identification of cues precisely related to personality perception can further improve the expressivity of synthetic speech [18].

The rest of the paper is organized as follows: Section 2 introduces personality and its measurement, Section 3 describes the approach proposed in this work, Section 4 reports on experiments and results, and Section 5 draws some conclusions.

2. PERSONALITY AND THE BIG FIVE

Personality is the latent construct accounting for "*individuals' characteristic patterns of thought, emotion, and behavior together with the psychological mechanisms - hidden or not - behind those patterns*" [6]. Among the different paradigms used to subsume personality (see [22] for an extensive survey), the one based on *traits*, i.e. on a limited number of dimensions accounting for consistencies in behavior, appears to be the more widely accepted. This applies in particular to the Big Five Factor Model [22], considered the "*latitude and longitude*" along which any other personality construct should be positioned [6]. For this reason, the personality assessments used in this work are based on the *Big Five Inventory* (BFI) [13], a questionnaire aimed at providing a score for each of the dimensions in the Big Five Model (extraversion, agreeableness, conscientiousness, neuroticism, openness).

2.1 The Big Five

An important empirical validation of the Big Five has been pointed out in the *lexical perspective* [14], one of the main theories of personality. The lexical perspective considers that the everyday language includes a large number of words describing personality traits (more than 15000 in English), but these can be grouped into a small number of clusters that actually correspond to the Big Five:

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

The main advantage of this phenomenon is that each personality can be described, at least in principle, with five scores accounting for how well the personality matches the words of each cluster (the list above shows some examples for each of the Big Five). Finding such scores is the goal of questionnaires like the BFI that are commonly applied in personality assessment.

2.2 Measuring Big Five Traits

There are several standard questionnaires aimed at scoring personalities along the dimensions corresponding to the Big Five (see [13] for a survey). The experiments of this work are based on the BFI version, called BFI-10, that includes only 10 of the original items of the BFI [13]. However these lead to results highly correlated with the full version of the BFI. The main advantage of the BFI-10 is that it allows one to perform personality assessments in a much shorter time than in the case of full BFI. After a judge has answered the 10 questions about a given person, the personality is described with five integer scores in the interval $[-4, 4]$. In order to make the assessment more reliable, each person is assessed by several judges and different scores are averaged. As a result, the final scores are distributed continuously in the same interval $[-4, 4]$.

3. AUTOMATIC PERSONALITY PERCEPTION

The personality perception approach proposed in this work includes three main steps: extraction of low level prosodic features, estimate of statistical features accounting for long-term variation of low level features, and mapping of statistical features into attributed traits.

3.1 Low-Level Feature Extraction

The low level features are pitch, formants, energy and speaking rate (measured indirectly through the length of voiced and unvoiced segments). These are not only the most important prosodic characteristics, but also the features most commonly applied in emotion recognition, a domain that has investigated a wide range of nonverbal behavioral cues in speech [20].

The low-level features are extracted, using PRAAT (version 5.1.15) [2], from 40 ms windows at regular time steps of 10 ms. The low level features are estimated on a frame by frame basis so they reflect only short term characteristics of vocal behavior. As personality perception is affected by long-term characteristics of vocal behavior, it is necessary to estimate the statistical properties of the low-level features over an entire speech clip.

3.2 Estimate of Statistical Features

The previous step extracts low level features every 10 ms, but the personality recognition is performed over an entire speech clip. Thus the low-level features cannot be used in their raw form, but rather through the estimation of their statistical properties. In this work we use, for each of the 6 low level features extracted at the previous step, the following four statistics: average, minimum, maximum and relative entropy. This last is a measure of uncertainty of a random variable. If X is a discrete random variable taking values in the set $\mathcal{X} = \{x_1, x_2, ..., x_n\}$, its entropy is :

$$H(X) = \frac{-\sum_{i=1}^{n} P(x_i) log(P(x_i))}{log(|X|)} \quad (1)$$

where $P(x_i)$ is the probability of $X = x_i$ and $|X|$ is the cardinality of $\mathcal{X}$. The term $log|X|$ works as a normalization factor, $H(X) = 1$ when the distribution is flat, while $H(X) = 0$ when only one value of $\mathcal{X}$ is represented. As the low-level features are 6 and the statistical features are 4, the total number of features extracted from each clip is 24.

3.3 Recognition

As described in Section 2, a personality assessment consists of 5 continuous values corresponding to the Big Five.

	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
1. This person is reserved	(1)	(2)	(3)	(4)	(5)
2. This person is generally trusting	(1)	(2)	(3)	(4)	(5)
3. This person tends to be lazy	(1)	(2)	(3)	(4)	(5)
4. This person is relaxed, handles stress well	(1)	(2)	(3)	(4)	(5)
5. This person has few artistic interests	(1)	(2)	(3)	(4)	(5)
6. This person is outgoing, sociable	(1)	(2)	(3)	(4)	(5)
7. This person tends to find fault with others	(1)	(2)	(3)	(4)	(5)
8. This person does a thorough job	(1)	(2)	(3)	(4)	(5)
9. This person gets nervous easily	(1)	(2)	(3)	(4)	(5)
10. This person has an active imagination	(1)	(2)	(3)	(4)	(5)

Table 1: 10-item questionnaire to measure the Big Five personality traits. The questions reported here correspond to those proposed in [13], but they are applied to a person different from the assessor rather than in a self-assessment form.

To perform a classification, the scores of each trait are split into two subsets called *High* and *Low*. The latter includes the samples that have a score lower than the average, while the former includes the samples for which the score is higher than the average. Each of the two subsets corresponds to a class. The classifier used in this study is a Support Vector machine (SVM) with RBF kernels [3].

4. EXPERIMENTS AND RESULTS

The experiments of this study have been performed over a corpus of 640 clips extracted from 96 news bulletins of *Radio Suisse Romande*, the French speaking Swiss national broadcasting service. Each clip includes only one speaker and the total duration of the corpus is around 7 hours. The average length of the clips is around 40 Seconds; 309 clips portray *journalists* and the other 331 portray *non-journalists*. The number of identities is 330 and the same identity is never represented in both training and test set. This is expected to ensure that what the system recognizes is the personality assessment and not simply the voice of the persons. A segment of 10 seconds has been randomly extracted from each of the clips and submitted to three judges for personality assessment. The goal is to reproduce a realistic scenario for spontaneous trait inference where the attribution of traits takes place in the first few seconds after an unknown individual is met. However, the prosodic features are extracted from the clips from where the 10 seconds long segments are extracted.

4.1 Personality Assessment

Each of three judges has assessed the whole corpus in separate sessions during which 30 clips had to be assessed. The order of the clips is random and it changes for each judge. The goal of this experimental setup is to avoid tiredness effects, i.e. to avoid the inevitable decrease of attention and concentration after a prolonged assessment effort.

The judges do not understand the language of the clips (French) and their mother tongue is Farsi. Furthermore, the clips do not contain any person or place name that could be recognized even by non-French speakers and have been selected so to be as neutral as possible in terms of content. This is expected to limit the effect of the verbal content on the perception of personality traits.

Each clip has been assessed by the three judges and has been assigned five scores corresponding to the dimensions of the Big Five model. The scores are the average of the scores assigned by each judge individually. For each dimension, the clips have been split into two groups: those who have a score lower than the average and those who have score higher than the average. The classes are called *Low* and *High* respectively.

4.2 Automatic Trait Attribution

The first experiment measures the effectiveness of the approach described in Section 3 in automatically recognizing whether a clip is in the *High* or *Low* class for each of the traits. An SVM with Radial Basis Function Kernel has been trained for each of the two classes using a k-fold approach: the entire dataset is split into k equal size subsets, $k-1$ subsets are used for training and the remaining one for testing. The procedure is repeated k times (each time, a different subset is used for testing) and the average performance of all k runs is reported as the overall performance of the classifier [1, 4]. In the experiments of this work, $k = 15$ and the performance is measured in terms of recognition rate (percentage of clips assigned to the correct class).

The results are reported in Table 2 and show that the recognition rate is significantly higher than chance for all of the dimensions except openness. However, the only dimensions for which the performance can be considered satisfactory are extraversion and conscientiousness. This is not surprising because certain traits become more evident while others more elusive depending on the data under examination [5].

However different assessors could have used the scales in different ways and the lack of results could be due to the low level of inter-rater agreement. Thus a second experiment has considered only the clips where the discrepancy between the different judges is lower than 3, this ensures that the inter-rater agreement is higher [7]. The approach is the same as the one described before (k-fold validation with $k = 15$) and the results are reported in Table 3.

While the performances of Tables 2 and 3 cannot be compared because they have been obtained over different datasets, the performance on openness is no significantly higher than chance and the noise due to inter-rater disagreement seems to be smoothed.

5. CONCLUSIONS

This paper has presented preliminary experiments on automatic personality trait attribution based on nonverbal vocal behavioral cues. The results show that the performance

Traits	Recognition Rate total	"High"	"Low"	Inter-rater Index
Extraversion	76.3	82.5	69.5	0.30
Agreeableness	63.0	75.9	47.7	0.30
Conscientiousness	72.0	77.0	65.8	0.32
Neuroticism	63.0	53.6	71.3	-0.11
Openness	57.9*	71.6	40.6*	-0.52

Table 2: Trait attribution performance. The results of this table have been obtained by using the average of the personality scores over the three assessors. When the difference with respect to a random classifier is not statistically significant, the values are denoted with a '*'. The negative inter-rater agreement values are due to the errors of one of the assessors that has misunderstood one of the assessors that has misunderstood one of the questions in roughly 15% of the questions.

Traits (Num. of clips)	Recognition Rate total	"High"	"Low"	Inter-rater Index
Extraversion(335)	79.4	84.8	73.8	0.78
Agreeableness(424)	64.7	62.2	66.9	0.52
Conscientiousness(423)	75.7	75.5	76.7	0.65
Neuroticism(360)	63.6	52.4	73.4	0.71
Openness(417)	62.8	75.2	47.7	0.35

Table 3: Performance over clips with higher inter-rater agreement.

is above chance at a significant level for all of the dimensions except openness, but the recognition rate is satisfactory only along two dimensions, namely extraversion and conscientiousness. However, the experiments of this work have to be considered preliminary and several details must be improved before reaching reliable conclusions. The first problem is that the nonverbal vocal features applied in this work are basic and important vocal characteristics, e.g. voice quality, have been neglected. Furthermore, the estimate of the entropy is based on the assumption that consecutive measures of the same prosodic feature are independent while this is clearly not the case.

Another important problem is that no normalization of the assessments has been applied across the raters and this introduces noise in the scores because different judges use the scales in different ways. Taking into account this effect can probably lead to more coherent scores and improve the results.

Both above problems will be addressed as a future work, in conjunction with the application of the same research approach to multimodal corpora where it will be possible not only to consider speech cues, but also nonverbal behaviors in faces, gestures, postures, etc.

Acknowledgments

This work is supported by the European Community's Seventh Framework Program (FP7/2007-2013), under grant agreement no.231287 (SSPNet), by the Swiss National Science Foundation through the Indo Swiss JRP "Cross-Cultural Personality Perception" (Grant 122936), and through the National Centre of Competence in Research on Interactive Multimodal Information Management (IM2). The authors wish to thank Olivier Bornet for building infrastructure for personality assessment.

6. REFERENCES

[1] C. M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.

[2] P. Boersma. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *in Proceedings of the Institute of Phonetic Sciences, Amsterdam*, volume 17, 1993.

[3] C. J.C. Burges. *A Tutorial on Support Vector Machines for Pattern Recognition*, 1998.

[4] K. J. Cios, W. Pedrycz, R. W. Swiniarski, and L. A. Kurgan. *Data mining: a knowledge discovery approach*. Springer, 2007.

[5] J.M. Dewaele and A. Furnham. Personality and speech production: a pilot study of second language learners. *Personality and Individual Differences*, 28(2), 2000.

[6] D.C. Funder. Personality. *Annual Review of Psychology*, 52:197–221, 2001.

[7] D.C. Howell. *Statistical methods for psychology*. Wadsworth Publishing Company, 2009.

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

[9] C.I. Nass and S. Brave. *Wired for speech*. MIT press, 2005.

[10] D.O. Olguin, P.A. Gloor, and A.S. Pentland. Capturing individual and group behavior with wearable sensors. In *Proceedings of the AAAI 2009 Spring Symposium*, 2009.

[11] M. Pantic and A. Vinciarelli. Implicit Human-Centered Tagging. *IEEE Signal Processing Magazine*, 26(6), 2009.

[12] F. Pianesi, N. Mana, and A. Cappelletti. Multimodal Recognition of Personality Traits in Social Interactions. In *Proc. of the 10th Int. Conf. on Multimodal Interfaces*, 2008.

[13] B. Rammstedt and O. P. John. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41:203–212, 2007.

[14] G. Saucier and L.R. Goldberg. The language of personality: lexical perspectives on the five-factor model. In J.S. Wiggins, editor, *The five-factor model of personality*. Guilford, 1996.

[15] K.R. Scherer. Personality markers in speech. In K.R. Scherer and H. Giles, editors, *Social markers in speech*. Cambridge University Press, 1979.

[16] K.R. Scherer and U. Scherer. Speech behavior and personality. In J. Darby, editor, *Speech evaluation in psychiatry*. Grune & Stratton, Incorporated, 1981.

[17] A. Todorov, A.N. Mandisodza, A. Goren, and C.C. Hall. Inferences of competence from faces predict election outcomes. *Science*, 308(5728):1623, 2005.

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

[19] J.S. Uleman, S.A. Saribay, and C.M. Gonzalez. Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, 59:329–360, 2008.

[20] A. Vinciarelli and G. Mohammadi. Towards a technology of nonverbal communication. In D. Gokcay and G. Yildirim, editors, *Affective Computing and Interaction: Psychological, Cognitive and Neuroscientific Perspectives*. IGI, 2010.

[21] A. Vinciarelli, M. Pantic, and H. Bourlard. Social Signal Processing: Survey of an emerging domain. *Image and Vision Computer Journal*, 27(12):1743–1759, 2009.

[22] J.S. Wiggins, editor. *The Five-Factor Model of Personality*. Guilford, 1996.