Automatic Chinese Personality Recognition Based on Prosodic Features

Huan Zhao, Zeying Yang, Zuo Chen, and Xixiang Zhang

School of Information Science and Engineering,
Hunan University, Changsha, Hunan, P.R. China 410082
chenzuo@hnu.edu.cn

Abstract. Many researches based on the English, French and German language have been done on the relationship between personality and speech with some relevant conclusions. Due to the difference between Chinese and other languages in pronunciation of acoustic characteristics, Chinese personalities and westerners, we put forward the Chinese and his personality prediction research in view. During the study, we collected 1936 speech pieces and their Big Five questionnaires from 78 Chinese. Built models for male and female with arguments of prosodic features such as pitch, intensity, formants and speak rate. Experiments' result shows: (1) the third formant has the same effect as the first two in prediction of personality; (2) combination of pitch, intensity, formants and speak rate as classification parameters can achieve higher classification accuracy(more than 80%) than in single prosodic feature.

Keywords: Automatic personality recognition, Chinese speech, Prosodic features, Personality traits, Big five.

1 Introduction

It is well known that personality is very important in our social activities. Effective assessment of personality is helpful in interpersonal communication, social relations, and career planning. As it is well known, every movement or single speech includes lot of information that can bespontaneous, unconscious, which largely references our personality[1]. Furthermore, our personality can also be referenced through our behavior, appearance[2], or even through videos[3] and audios[4].

In this work, we focus on the speech and personality. A lot of work have already been done on this, with good results obtained. For example, In article [5]and[6], speech audio taken from a specialist have been classified in Big Five based on acoustics and prosodic features. In[7], the relationship between prosodic features and personality were explored. The authors used pitch, energy, first two formants, voiced-time and unvoiced-time as prosodic features. A corpus of 640 speeches was built which is about 10 seconds in average, and personality traits grades was assessed by listeners. Then they used Logistic regression and Support Vector Machines(SVM) to classify the data. Following the procedure described above, they claimed an accuracy of about 70%.

X. He et al. (Eds.): MMM 2015, Part I, LNCS 8935, pp. 180–190, 2015.

Previous works that studied the relationship between prosodic features and personality used a corpus of either English, French or German but not Chinese. When collecting data, they mainly took the thirds assessment as the standard value in personality traits perception. As discussed in [8], there are some differences between Chinese pronunciation and western language (such as English), so was the Chinese personality and western's. Therefore, we put forward the Chinese personality research based on speeches. One of the most important reason this work is proposed is that past researches are mostly using the third person's(listeners) perception results as the prediction standard values in perception of speakers' personality. As we know, the third person's perception results may not rightly response speakers' personality if he is not familiar with the speaker. In order to be able to accurately predict the speaker's personality characteristics, we adopt the speaker to self-assessment as the standard to determine the classification accuracy of the model. In view of the above points, our mainly work is to: (1)Collecting 78 persons' speeches and their own Big Five questionnaire; (2)Taking pitch, energy, first three formants, voiced-time and unvoiced-time as prosodic features to test the effect of personality recognition; (3)Building personality predictions' model for male and female, respectively using an SVM.

The rest of this paper introduces theories and principles of our assumption, experiments and results. At the end of the paper, we make a summary and some prospects about future works.

2 Speech and Personality

2.1 Assessment of Personality

Personality is *individual's characteristic patterns of thought, emotion, and behavior together with the psychological mechanisms-hidden or not-behind those patterns*[9]. Many experts built lots of personality assessment models for the importance of personality traits. The famous models are: Big Five[10], Sixteen Personality Factor Questionnaire(16PF)[11], Eysencks personality theory[12] and so on. In these models, Big Five has wide applicability[13] and high credibility[14] than other models. So we take it as personality standard. Big Five is the result of vocabulary method research. Researches found that personality can be covered just by the following 5 traits:

Extraversion(Ex): high extraversion means more enthusiastic, sociable, energetic, optimistic and adventurous etc.

Neuroticism(Ne): includes characteristics such as: anxiety, hostility, repression, impulsion and frailty etc.

Openness(Op): people has high score in this trait will be more imaginative, wisdom, aesthetic, creativity and affettuoso and vice versa.

Agreeableness(Ag): features are trust, altruism, frank and modesty etc.

Conscientiousness(Co): people with high score in conscientiousness indicates he/she is fairer, more dependable, self-disciplined and dutiful than lowers.

Questionnaire of Big-Five that we use in this work includes 60 questions. Each question has five answers: "strongly disagree", "disagree", "neutral", "agree" and "strongly agree", which can be transformed into scores from 1 to 5. As every personality trait owns 12 questions in this inventory, every trait's total score is ranging from 12 to 60.

2.2 Speech and Personality

In 1927, Edward Sapair presented that there is some relationship between personality and voices[15]. Then papers about using vocal behavior to percept personality have sprung up, and most of them are based on prosodic features and personality traits.

Prosodic features mainly include pitch, energy and rate. Pitch is used to indicate the ears' feeling about vibration of objects and is decided by frequency of sound wave. Generally, the higher the frequency of the vibration, the higher pitches; so is the lower vibration frequency. The energy of voice is measured by intensity, which can express average energy of voice. Voice time interval can indicate speak rate.

In previous researches, some of them take a kind of prosodic features as characteristic to explore. In[16], pitch is taken as the parameter to percept personality and the result indicates that a higher pitch represents a more active personality for him or a more exoscopic personality for her. But most of previous researches take two or more prosodic features as target. In [17], five assumptions are used to explore the relationship between competence and benevolence of personality and prosodic features. And its conclusion shows that high speed and high pitch suggests high competence and vice versa. But high speed and low pitch variation indicates low benevolence ratings. Many papers have explored topic of competence and prosodic features[7][18] and most of them got a simple result: higher speed leads to higher competence ratings. Moreover, in some other articles, voices volume[19], tone quality and speech's pause time[20] have also been taken in to consideration.

The above researches mainly concentrated in the late of the 20th century. In recent years, the study of personality and prosodic features took more characteristics in consideration. In [21][22], authors take pitch spectrum, pitch grade, rate and intensity as prosodic features. In the aspect of personality, self-assessment was adopted to evaluate the sincerity, excitability, ability, etc. In a research about excitability[23], the parameters include not only pitch spectrum, pitch variation range, rate, but also eye movement, blinking and eyebrow movement.

As mentioned before, people made a lot of research on the relationship between the personality and speech, and achieved rich results. But all of these studies used corpus based on English, German, French or other western language and their objects are westerns too. In[8], authors discuss the pronunciation difference between Chinese and English. So due to the difference between Chinese and other western language in pronunciation, in this work, we try to use pitch, energy and rate as prosodic features to predict Chinese's personality.

3 Approach

The APR(Automatic Personality Recognition) approach proposed in this article mainly includes four steps: (1)collecting corpus for the experiments; (2)extracting low-level short-term and estimate long-term prosodic features from the speeches; (3)matching prosodic parameter and personality traits; (4)building statistical model of personality traits based on prosodic characteristics and test its performance.

3.1 Building Corpus

As the limit of corpus, we have to build a corpora for experiments. There are some principles in collecting speech pieces: firstly, subjects' mother language must be Chinese, secondly, they should have complete language ability, and can use Chinese to express their own ideas and things fluently.

3.2 Extracting Parameters

After collecting the data, we can extract features. This can be separated into 2 steps: short-term features extraction and long-term features extraction. As long-terms used in this experiment can be counted from short-terms, the accuracy of short-term is very important.

In this work, the short-term features extracted are first three formants, intensity, pitch, voiced time and unvoiced time. The formants, which are *"the spectral peaks of the sound spectrum of the voice"*[23] was defined by Gunnar Fant. Intensity is used to describe energy of the speech clips. Pitch represents the main correlation between tone and intonation. Pitch, rate and energy are the most important characteristics of prosody, so these four features are used in our experiments[7].

To extract these features, we used a wildly applied software–Praat[24]. The features are extracted from 25ms analysis windows at step of 10ms. Every clip's feature-parameters can be stored in a matrix $F = (\boldsymbol{f_0}, \boldsymbol{f_1}, \cdots, \boldsymbol{f_8})$, the first row($\boldsymbol{f_0}$) represents time(from 0 to end of the clip), vectors from $\boldsymbol{f_1}$ to $\boldsymbol{f_8}$ correspond to pitch, intensity and first three formants and its bandwidth. After getting these three features, we can separate voiced and unvoiced time bites from the speeches and stored in matrix $T = (\boldsymbol{t_0}, \boldsymbol{t_1})$.

The long-term features can be estimated from the data got above. Maximum, minimum and average are used to describe the dynamic range, and standard deviation is a measure of data average dispersion degree. These four statistical features are enough for pitch, intensity and formants, but for time we add another dimensionality: voiced-time/unvoiced-time.

So far, we have built our data set which includes two matrix(one for male, another for female) like this $D = (\boldsymbol{d_0}, \cdots, \boldsymbol{d_n})$, n=33. $\boldsymbol{d_0}$ to $\boldsymbol{d_4}$ indicate the classes of personality traits. Each trait is separated into three classes; $\boldsymbol{d_5}$ to $\boldsymbol{d_8}$ belong to pitch features; $\boldsymbol{d_9}$ to $\boldsymbol{d_{12}}$ are the intensity parameters; $\boldsymbol{d_{13}}$ to $\boldsymbol{d_{24}}$ are formant parameters; the last 9 vectors are time parameters.

3.3 Studying and Classifying

In the experiments, Support Vector Machines(SVM)–a binary classifier was used to train and classify the corpus. As we have three classes, is necessary to improve original SVM. So first, we use the one-vs-all method to translate three classes into binary, then use SVM again.

The purpose of SVM is finding a separating hyperplane $\boldsymbol{w}^T\boldsymbol{x} + b$ so that the points belong to different classes can space on the either side of hyperplane. The points closest to the hyperplane is called support vectors and the distance from them to separating hyperplane $|\boldsymbol{w}^T\boldsymbol{P} + b|/||\boldsymbol{w}||$ must be the maximum. Therefore the goal is find w and b values in the classifier,we can write it as:

$$max_{w,b}\{min_n(sign \cdot (\boldsymbol{w}^T\boldsymbol{x} + b)) \cdot \frac{1}{||\boldsymbol{w}||}\} \tag{1}$$

It's difficult to solve this problem directly, unless we convert it to another form. A method called Lagrange multipliers has be widely used to solve this type problems. Hence, the function above can be writen as:

$$max_\alpha[\sum_{i=1}^{m}\alpha - \frac{1}{2}\sum_{i,j=1}^{m} sign^{(i)} \cdot sign^{(j)} \cdot a_i \cdot a_j\langle x^{(i)}, x^{(j)}\rangle] \tag{2}$$

with constraints:

$$\alpha \geq 0 \quad and \quad \sum_{i=1}^{m}\alpha_i \cdot sign^{(i)} = 0 \tag{3}$$

Hereto, the main work of SVM is solving these α[25][26]. There are many methods that can solve this optimization problem. But to train our data, the Sequential Minimal Optimization(SMO), which is published by John Platt in 1996[27] is been chosen. *"SMO breaks this large quadratic programming(QP) problem into a series of smallest possible QP problems. These small QP problems are solved analytically, which avoids using a time-consuming numerical QP optimization as an inner loop"*. Thus, SMO is one of the fastest method used to solve these kind of problems. So using it in our experiment for training our model is best.

k-fold cross validation[28] is also used in the experiment. We split the data D into 10(as stated in[28] *"10 folds can get a better than the more expensive leave-one out cross validation"*) equal subsets $D = \{D_1, D_2, \cdots, D_{10}\}$ at random, every time $D_i(i \in \{1, 2, \cdots, 10\})$ is used for test, and the other 9 subsets are used to train the model. So each fold of training and classifying is independent and the end result is more convincing.

4 Experiments and Results

In this section, experiments and its results are mainly described in three subsections: (1)data collection; (2)personality traits recognition performance comparison when taken first three formants as features; (3)personality traits recognition performance comparison when taking pith, intensity, formants and time as prosodic features. In (2) and (3), correctly classified rate was used to measure prosodic features performance in personality traits recognition.

4.1 Data Collection Results

As mentioned before, we interviewed 78 Chinese, whose mother language are Chinese, and whose are able to communicate in Chinese in general situations. Of all the subjects, there are 38 males and the others are females. For each of subject, firstly we got his(or her) NEO-FFI which contains 60 questions. After that, we recorded their speech clips which are spoken in dispassionate by letting him(or her) describe some thing or just answer specific questions. About 25 dispassionate clips are taken from every participant and most clips are in 10s to 15s. So finally the corpus was composed by 1936 speech clips 975 bites belonging to males and the other 961 chips belonging to females, as shown in Fig. 1.

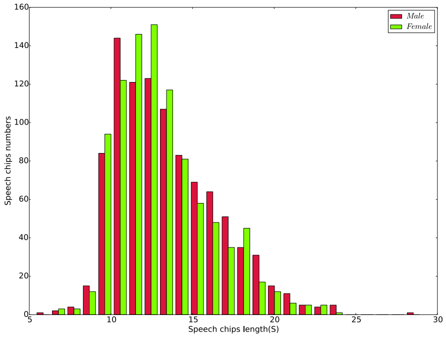

Fig. 1. Distribution of speech clips length. In this chart, male's mean length of speech clips is 13.455s, and female's is 13.158s. The number of speech clips, whose length between 9s-15s is 682 and 731 for male and female, respectively.

Figure 2 shows the NEO-FFI scores of each subject. For Ag and Ex, their score distribution is around the average 36; and the other three traits' average score is about 40. Personality is hard to measure in quantitative because of it selfs characteristic, and not needed. Therefore in the experiments below, personality traits will be divided into three grades based on the traits score.

4.2 Formants Predict Personality

Formant is a prosodic feature, which refers to vibration frequency of pronunciation and intonation. People were aware of the importance of first two formants, but the importance of the third formant in phonetics were enlightened in papers such as [7] and[29]. In this work, the relationship between personality traits were judged through classification results of men's and women's different traits based on first three formants, which are taken as prosodic features.

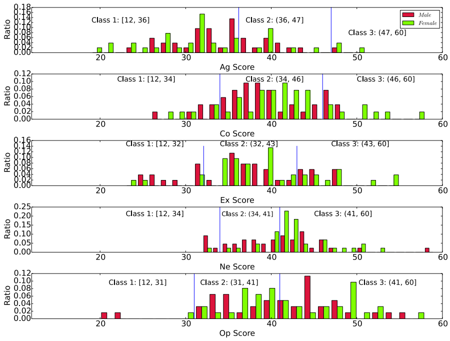

Fig. 2. NEO-FFI score distribution

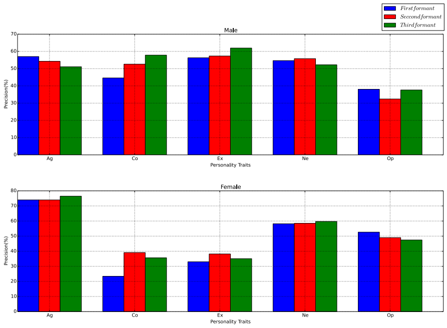

Fig. 3. The personality traits classification result based on single formant

For two subgraphs in figure 3, the x axis represents the distribution of each personality characteristics, the y axis represents classification accuracy. In every characteristic, the three rectangles stand for the first three formants classification accuracy. For male, we get a best accuracy of 61.95% for Ex with is obtained by using the third formant. But using the first formant we get an accuracy of 56.31% and an accuracy of 57.33% for the second formant. These accuracy results means Ex is easy to classify. The lowest recognition accuracy is from Op trait. Its three formants classification accuracy are only 38.05%, 32.41% and 37.64%, respectively. In general for the two personality traits Ag and Op, the first formant classification effect is the best; For personality traits Co and Ex, the third formant has the highest classification accuracy. For the other personality trait the second formant effect is slightly better than the other two. For female, the best accuracy result is obtained from Ag, which accuracy rates are 73.98%, 73.99%, 76.48%. That is about 20% higher than the accuracy rates obtained for male. The Co and Ex have the poor effect in classification accuracy, which are both less than 35%. In the picture, although each single feature recognition

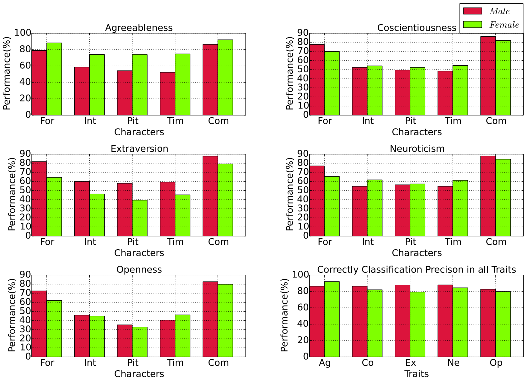

Fig. 4. Personality recognition results based on prosodic features(Formants(For), Intensity(Int), Pitch(Pit), Time(Tim), Combination(Com))

rate is low, it can be known that any formant can play similar role when used as prosodic features in personality traits recognitions. Although the role of the third formant in the phonetics is not well-known, we can apply the third formant to personality traits recognition based on prosodic features.

4.3 Prosodic Predicts Personality

In the above section, we explore the formant's influence in personality traits. Pitch, sound intensity, voiced time, unvoiced time, the first three formants, and the combinations of them are used to explore the impact on the classification of personality traits. As shown in figure 4, the first 5 subgraphs represents 5 different personality traits. The abscissas of the 5 subgraphs represent formant, intensity, pitch, timing and the combination of them. The ordinates are classification accuracy rate. For different personality traits, prosodic features as the basis of classification results in different identification accuracy rate. In general, when using single prosodic feature, formant can get better effect than other threes, especially for Op. Also for Op no matter which single prosodic feature(or their combination), its recognition rate is lower than the other four personality traits. As shown in the figures, either male or female, the recognition rates when based on single prosodic feature, are lower than the recognition rates obtained by combining those features. For Ag, when using a single feature, the average recognition rate of men and women are 58.1% and 74.71% respectively; 84.62% and 87.62% in combination. In each other personality traits, also it can be seen that combined characteristics receive a higher classification accuracy rate than a single characteristic with a gain of about 20% accuracy. This means each prosodic features can reflect one or more aspects of our personality.

When it comes to classification, there are some mistakes more or less. The confusion matrix(see table 1) will show classification more directly and concretely when using the combined prosodic features. The table is divided into two parts,

Table 1. Combination characteristics classification results' confusion matrix(%)

Traits		Male			Female		
		1	2	3	1	2	3
Ag	1	96.00	0	4.00	87.33	12.67	0
	2	4.22	86.89	8.89	0	77.20	22.80
	3	6.95	11.79	81.26	0	8.72	91.28
Co	1	97.00	1.00	2.00	92.00	4.00	4.00
	2	13.43	77.14	9.43	9.07	76.17	14.77
	3	5.14	10.29	84.57	6.80	16.80	76.40
Ex	1	92.00	5.00	3.00	90.67	5.33	4.00
	2	6.78	89.04	4.17	9.87	77.07	13.06
	3	11.33	12.34	76.33	9.56	14.09	76.32
Ne	1	89.71	8.29	2.00	91.82	7.27	0.91
	2	15.40	83.40	1.20	31.33	67.34	1.33
	3	19.20	15.20	65.60	30.63	8.12	61.26
Op	1	81.60	10.80	7.60	85.00	8.00	7.00
	2	15.11	80.22	4.67	10.96	80.43	8.61
	3	17.09	11.64	71.28	12.57	20.86	66.57

men and female, and then divided into the five personality traits(Ag, Co, Ex, Ne, Op) according to Big Five personality model. Each of the personality trait is partitioned into three categories according to the NEO-FFI score. Every value in the table represents the classification accuracy. For example, the value 96% in third row and third column is said that for male's Ag trait, when using the model trained in SVM to classify speeches belonging class one, 96% of them are identified correctly. And the remaining 4% are classified as class three incorrectly, and so on. For each personality trait, their recognition rates are showing a phenomenon that the first class performs best, the second class performs better than the third.

5 Conclusions

The main work of this paper is the recognition of personality in Chinese speeches prosodic features. We use the three main prosodic features: pitch, energy and speed rate with the personality traits, which are got from NEO-FFI. After data collection, the first three formants have been analyzed for personality recognition. Even though the importance of the third formant is still unclear, one of the finding of this paper is that the third formant has the same effect in automatic personality recognition as the first two. In another personality traits recognition experiment, a comparison between single character and combination characters has been made. This experiment shows that combination's accuracy rate is higher than the singles in average by 20%. Also by combining prosodic features for, we classification can get a better recognition rate. As other papers are based on different data set, comparing our work with theirs can hardly get convincing.

As mentioned before, we present our personality traits not only in speech but also in many other ways. So exploring prosodic features with other characteristics such as facial movements, body language[30] even eye movements is very meaningful. As China is a very large country, the research about automatic personality recognition based on the local language is also meaningful.

References

[1] Barbero, C., Zovo, P.D., Gobbi, B.: A flexible context aware reasoning approach for iot applications. In: 2011 12th IEEE International Conference on Mobile Data Management (MDM), vol. 1, pp. 266–275. IEEE (2011)

[2] Uleman, J.S., Saribay, S.A., Gonzalez, C.M.: Spontaneous inferences, implicit impressions, and implicit theories. Annu. Rev. Psychol. 59, 329–360 (2008)

[3] Mayer, R.E.: Multimedia learning. Cambridge University Press (2009)

[4] Olivola, C.Y., Todorov, A.: Elected in 100 milliseconds: Appearance-based trait inferences and voting. Journal of Nonverbal Behavior 34(2), 83–110 (2010)

[5] Polzehl, T., Moller, S., Metze, F.: Automatically assessing personality from speech. In: 2010 IEEE Fourth International Conference on Semantic Computing (ICSC), pp. 134–140. IEEE (2010)

[6] Polzehl, T., Moller, S., Metze, F.: Automatically assessing acoustic manifestations of personality in speech. In: 2010 IEEE Spoken Language Technology Workshop (SLT), pp. 7–12. IEEE (2010)

[7] Mohammadi, G., Vinciarelli, A.: Automatic personality perception: Prediction of trait attribution based on prosodic features. IEEE Transactions on Affective Computing 3(3), 273–284 (2012)

[8] Haiying Li, Y.W.: Comparative study on the phonetic features chinese, english and japanese. Social Science Forum 9, 176–180 (2009)

[9] Allport, G.W.: Personality: A psychological interpretation (1937)

[10] Komarraju, M., Karau, S.J., Schmeck, R.R., Avdic, A.: The big five personality traits, learning styles, and academic achievement. Personality and Individual Differences 51(4), 472–477 (2011)

[11] Cattell, R.B., Eber, H.: Sixteen personality factor questionnaire (16pf). Institute for Personality and Ability Testing, Champaign, Illinois, USA (1972)

[12] Eysenck, S.B., Eysenck, H.J., Barrett, P.: A revised version of the psychoticism scale. Personality and Individual Differences 6(1), 21–29 (1985)

[13] Hattie, J.: Visible learning: A synthesis of over 800 meta-analyses relating to achievement. Routledge (2013)

[14] John, O.P., Naumann, L.P., Soto, C.J.: Paradigm shift to the integrative big five trait taxonomy. Handbook of Personality: Theory and Research 3, 114–158 (2008)

[15] Sapir, E.: Speech as a personality trait. American Journal of Sociology, 892–905 (1927)

[16] Argyle, M.: Bodily communication. Routledge (2013)

[17] Ray, G.B.: Vocally cued personality prototypes: An implicit personality theory approach. Communications Monographs 53(3), 266–276 (1986)

[18] Collier, G.J.: Emotional expression. Psychology Press (2014)

[19] Lindzey, G., Gilbert, D., Fiske, S.T.: The handbook of social psychology. Oxford University Press (2003)

[20] Burgoon, J.K., Guerrero, L.K., Floyd, K.: Nonverbal communication. Allyn & Bacon, Boston (2010)

[21] Schmitz, M., Krüger, A., Schmidt, S.: Modelling personality in voices of talking products through prosodic parameters. In: Proceedings of the 12th International Conference on Intelligent User Interfaces, pp. 313–316. ACM (2007)

[22] Trouvain, J., Schmidt, S., Schröder, M., Schmitz, M., Barry, W.J.: Modelling personality features by changing prosody in synthetic speech (2008)

[23] Fant, G.: Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations, vol. 2. Walter de Gruyter (1971)

[24] Boersma, P.: Praat, a system for doing phonetics by computer. Glot International 5(9/10), 341–345 (2002)

[25] Bishop, C.M., et al.: Pattern recognition and machine learning, vol. 1. Springer, New York (2006)

[26] Harrington, P.: Machine Learning in Action. Manning Publications Co. (2012)

[27] Platt, J., et al.: Sequential minimal optimization: A fast algorithm for training support vector machines (1998)

[28] Kohavi, R., et al.: A study of cross-validation and bootstrap for accuracy estimation and model selection. IJCAI 14, 1137–1145 (1995)

[29] Mohammadi, G., Vinciarelli, A., Mortillaro, M.: The voice of personality: mapping nonverbal vocal behavior into trait attributions. In: Proceedings of the 2nd International Workshop on Social Signal Processing, pp. 17–20. ACM (2010)

[30] Pianesi, F., Mana, N., Cappelletti, A., Lepri, B., Zancanaro, M.: Multimodal recognition of personality traits in social interactions. In: Proceedings of the 10th International Conference on Multimodal Interfaces, pp. 53–60. ACM (2008)