EMPIRICAL ARTICLE

Reading and Self-Presentation Speech Acoustic Analysis for Identification of Personality Traits

Anastasia S. Panfilova [*a], *Nikita A. Pospelov*[b], *Denis V. Parkhomenko*[c], *Ekaterina A. Valueva*[a]

[a] *Institute of Psychology Russian Academy of Sciences, Moscow, Russia*
[b] *Institute for Advanced Brain Studies of Moscow State University, Moscow, Russia*
[c] *Huawei Technologies, Moscow, Russia*

Abstract*. The research is devoted to the diagnostics of the person's psychological properties by means of the voice acoustic characteristics analysis. The research is carried out on the example of psychodiagnostics of Big Five traits and Circumplex of Personality Metatraits, as well as the level of Crystallized and Fluid types of intelligence. It was demonstrated that the use of different scenarios, experimental situations and formulated tasks can increase the effectiveness of diagnosing a number of traits. This goal was achieved by creating sets of data taking into account the acoustic characteristics of the examinee when reading texts of two types: text 1 with a neutral emotional tone (Stanislav Lem "Solaris") and text 2 with a negative tone about sufferings of people during the blockade of Leningrad ("Memories of Lihachev") as well as conducting interviews with a simulated situation of employment with a set of questions. The study involved 356 subjects whose voices were recorded while reading texts of two types and answering 12 questions of audio-interviews. We found that the Conscientiousness trait was best diagnosed in males by text reading 1 (ROC-AUC = 0.76), and in females by interview questions (ROC-AUC=0.75). Traits related to emotional stability and mental health (GM, GP) are also best diagnosed in both men and women by text reading 1. An increase in the diagnostic accuracy of Crystallized intelligence in men was shown when using acoustic voice characteristics in text reading 1 (ROC-AUC=0.7).*

Keywords: *Crystallized intelligence, Fluid intelligence, Personality computing, Personality traits, Social signal processing*

Research on the relationship between psychological characteristics and characteristics of the voice has a history of nearly a century [26]. As Sapir notes "There is one thing that strikes us as interesting about speech: on the one hand, we find it difficult to analyze; on the other hand, we are very much guided by it in our actual experience. …none is entirely lacking in the ability to gather and be guided by speech impressions in the intuitive exploration of personality" [26]. Because of the notion that atomistic analysis makes no sense [1]. and due to

the lack of technical means, the first studies of the relationship between speech and psychological characteristics in the early 20th century were conducted based on an impression of the speaker's voice "in general": subjects were asked to listen to audio recordings and evaluate various physical (age, height, build, etc.), social (e.g., profession, political views) and psychological (extraversion, dominance, etc.) characteristics of the speaker.

Early work has shown that some voice characteristics can be related to personality

traits. In particular, Mallory and Miller [16] found weak but statistically significant correlations of introversion with high pitch, inadequate loudness, lack of resonance, and unconfident manner of speech. Other studies have demonstrated that extraversion and introversion are related to the pace of speech [10, 25].

Current studies, based mainly on the Big Five model of personality, support and broaden previous data. A study by Park *et al.* [22] demonstrated that extroverts, compared to introverts, had shorter pauses before answering questions. In Mairesse *et al.* [15] self-reported extraversion was shown to correlate with speech rate. Biel *et al.* [4] data suggested that extraversion could be predicted by longer speaking time and decreased number of pauses. Stern *et al.* [30] conducted a large secondary data analysis combining eleven independent datasets (2217 participants). They found that self-reported extraversion, dominance, and openness to experience had negative relationship with voice pitch, neuroticism had a positive one, and that there were no correlations between personality traits and mean formant position.

It is important to note that the prediction accuracy of personality traits measured by self-report and expert methods may differ. Thus, although some studies found negative correlations between extraversion and voice pitch [16, 30], others (e.g. [15; 4] found inverse relationships using observer's ratings as personality traits measures. In Mairesse *et al.* [15] the prosodic markers for both observed and self-reported extraversion were intensity variability and mean intensity. On the other hand, emotional stability as measured by self-reported extroversion was characterized by low intensity variability and low mean intensity, whereas these vocal properties did not play a role in external observer assessments. The authors hypothesized that the model for determining personality traits should switch from evaluations by external observers to self-report evaluations, because traits with high obviousness (extraversion) are more accurately evaluated by external observers, whereas traits with low obviousness (emotional stability) are more accurately evaluated by self-report, and Polzehl *et al.* [23, 24]. found that pitch range, speech rate, intensity, loudness, formants, or spectra can predict Big Five elements. However, these and more recent works used expert assessments

of speakers' psychological traits. The authors predominantly used regression analysis and SVM (Support Vector Machine). With the development of neural network methods, research using multilayer perceptron (MLP), supplemented by a model for analyzing the verbal side of speech (LSTM) have appeared [2]. In this paper, the maximum classification quality (proportion of correct classifications) achieved was as follows: openness to experience (77%), conscientiousness (63%), extraversion (64%), Agreeableness (61%), and neuroticism (68%). The method proposed by Carbonneau [7], which relies on the use of spectrograms and SVMs, increased the recognition efficiency of Agreeableness to 65% and neuroticism to 70%, while decreasing the prediction quality of the other indicators. Further works extended approaches to the characteristics selection and performed comparison of different neural network models. E.g., Tayarani *et al.* [33]. proposed to use the analysis of the pause fillers ("ehm", "uhm") in speech. When comparing *Cascade Forward Neural Network* (CFNN), *Feed Forward Neural Networks* (FFNN), *Fuzzy Neural Networks* (FNN), *Generalized Regression Neural Networks* (GRNN), k *Nearest Neighbors* (kNN), *Linear Discriminant Analysis* (LDA), *Naive Bayes Classifier* (NB), *Support Vector Machines* (SVM) and using the PCA-QEA approach to feature selection, the LDA classifier was shown to provide significant increase in classification quality for experience openness and extraversion. In terms of analyzing the frequency of the predictors selected by the algorithm, a lower frequency of delta coefficient selection should be note. One possible explanation is the fact that indicators should capture temporal variations, but pause fillers tend to be pronounced as long vowels, in which speech properties remain stable and, therefore, no major changes are observed. The main exceptions to this general pattern were observed for extraversion, where delta regression coefficients were chosen more frequently in the RMS and basic tone frequency (F0) groups. The next result was that the first two small-frequency cepstral coefficients (MFCC) were selected more frequently for conscientiousness and neuroticism.

In terms of experimental design, the Guidi *et al.* [12] study can be highlighted. The subjects were asked to read the text "The Universal Declaration of Human Rights" twice before and after the experiment, for three minutes. The

subjects were asked to comment on a set of images from the Thematic Apperceptive Test between the readings of the text. The subjects also completed the Spielberger Anxiety Test. No model was constructed in this study, but a correlation analysis was conducted. It was shown that the mean values of the acoustic measures of the two text readings were negatively correlated with the evaluation of the "Communicativeness" parameter; significant estimates of correlations with other measured personality traits were also found.

Current trends in personality traits diagnostics include the use of deep learning methods and the combination of both verbal and nonverbal speech components analysis along with video analysis to assessing the dynamics of emotional state [19].

The data

The final sample of Russian-speaking subjects who completed the tasks was 356, including 257 females (mean age 34.8) and 99 males (mean age 30.4) [21]. Data collection of psychodiagnostic data and audio-interview recordings was performed using the developed Internet platform without any special organized conditions for voice recording. Due to the large volume of tasks, the subjects were allowed to take the study in several stages. A total of 5,701 audio recordings were obtained (4,134 for women and 1,567 for men).

A. Psychological diagnostics

The study determined the following psychological characteristics of the subjects: disposition of basic personality traits, verbal intelligence, nonverbal intelligence.
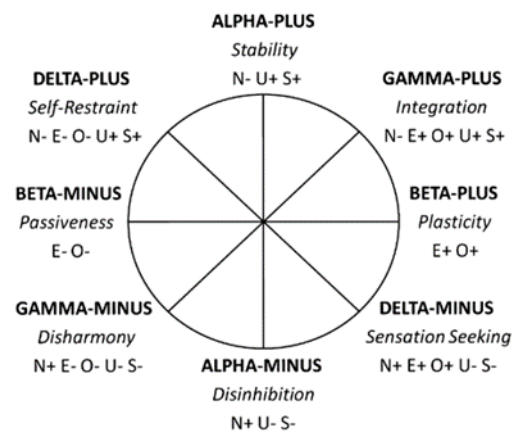
1) Personality traits

The Big Five model [8] is the most popular model of personality in psychology. It postulates that a variety of person's thoughts, feelings, and behaviors could be mapped into five broad dimensions (factors): Openness to experience (O), Conscientiousness (C), Extraversion (E), Agreeableness (A), and Neuroticism (N).

In our study, we used the Big Five Inventory (BFI; John *et al.* [13]) in Russian adaptation [28]. The BFI consists of 44 items aimed at measuring five main domains of the Big Five model: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness.

Despite the predominant role of the Big Five model, it is not free from criticism [5]. The main concern arises from intercorrelations usually

found between five personality traits [11]. A hierarchical structure and the existence of higher-order personality factors are suggested instead [9]. A related model, the Circumplex of Personality Metatraits (CPM), postulates the existence of two orthogonal metatraits (Alpha/Stability and Beta/Plasticity), with another metatrait representing General Personality Factor (Gamma/Integration), and Delta/Self-Restraint metatrait which is the combination of high stability and low plasticity (or vice versa). The positive and negative poles of each metatrait are defined separately and can be represented by specific combinations of the Big Five traits (see Fig. 1). For example, Alpha-Plus is characterized by low Neuroticism, high Agreeableness, and high Conscientiousness, whereas Delta-Minus includes high Neuroticism, Extraversion, and Openness combined with low Agreeableness and Conscientiousness [31]. We have used the Russian version of The Circumplex of Personality Metatraits Questionnaire [32] which consists of 72 items intended to measure each of the eight metatraits.

Figure 1. Circumplex of Personality Metatraits. N = Neuroticism; E = Extraversion; O = Openness to Experience; U = Agreeableness; S = Conscientiousness; + means a positive pole of the trait; - means a negative pole of the trait. From (Strus & Cieciuch, 2017). Copyright 2016 by Elsevier Inc.

The answers to the questions obtained by both techniques were presented in the Likert scale, which assumes the tested person should express the degree of agreement and disagreement with the statement on a scale of 1 to 5, where 1 = strong disagreement, and 5 = strong agreement. The resulting distributions of scores are presented in Fig. 2.
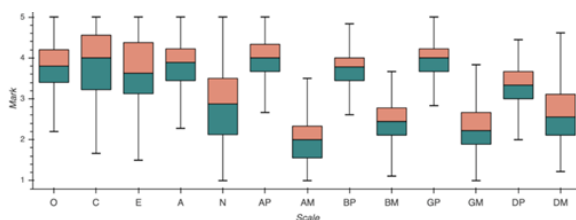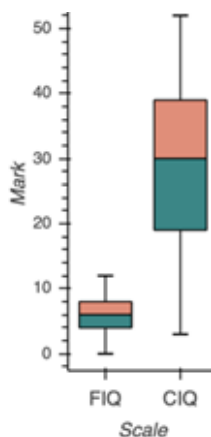
Figure 2. The results of the subjects' scores for different psychodiagnostic variables.

2) Crystallized intelligence (CIQ)

Crystallized intelligence is the ability to reason based on previously acquired knowledge. It is usually measured by verbal tasks involving vocabulary, reading comprehension, analogies, etc. We used three verbal scales in Russian: analogies (20 items for 6 min), generalization (20 items for 7 min) [36] and deduction (16 items for 8 min) [3]. The overall measure of crystallized intelligence was computed as a sum of scores for individual scales.

3) Fluid intelligence (FIQ)

Short form of the Raven's Advanced Progressive Matrices was used as a measure of fluid intelligence [6]. It consists of twelve 3×3 matrices of geometric shapes with one missing item that should be found among eight alternatives. This test is intended to measure the core of fluid intelligence - inductive reasoning and analytical thinking ability. Fig. 3 shows the distribution of tested subjects' results for the two techniques that measure intellectual ability.

Figure 3. Scatter diagram of the results of the Advanced Progressive Matrices (raven) and Intelligence structure test (verb).

B. Reading and audio-interview

To test the hypothesis of the study, the subjects were given 2 text fragments to read, while their voice recording was made. The first block (Text 1) was a fragment of the science

fiction novel "Solaris" by Stanislaw Lem (average reading time 116 seconds) - it was an emotionally neutral text. The second fragment (Text 2) was taken from the memoirs of D.S. Likhachev, describing the besieged Leningrad during World War II (average reading time 118 seconds). This text was heavily emotionally loaded, as it describes scenes of suffering, hunger, and death.

Next, the subjects were asked to imagine themselves in a situation of employment and an audio interview. In this regard, the respondents had to answer 12 questions, which were recorded on the audio by a female voice with a neutral intonation.

Interview questions:

1. Introduce yourself, please.
2. You have two minutes to briefly tell the most important things about yourself.
3. What kind of manager would you not work with?
4. What are your strengths?
5. Name your two shortcomings, describe in detail what you mean by that.
6. By what criteria did you choose where to study after high school?
7. Were you interested in learning?
8. Tell us, please, what exactly have you been doing in the last 2 years for your development, learning on your own initiative?
9. Please tell us about your favorite thing you like to do.
10. What activities you don't like?
11. Tell us about your accomplishment.
12. What do you consider as your failure?

The delta coefficients (numerical derivatives) were additionally computed for each of these descriptors (F0, ZCR, RMS, MFCC).

The proposed approach

In order to determine variations in the voice of an individual person it is necessary to analyze his speech features in the recording, where personality traits are least manifested.

In Section A we describe the variants of combining the two recordings (text reading and answering a question), as well as the use of averaged data on the examinees' answers.

In Section B we describe the procedure for selecting the basic model to test the hypothesis

of this study.

In Section C we provide the psychological context of the datasets used in this study.

A. Additional dataset

Let I be the matrix of acoustic characteristics values obtained from the subject's answers:

$$I = \begin{pmatrix} a_{11} & ... & a_{1i} \\ ... & ... & ... \\ a_{n1} & ... & a_{ni} \end{pmatrix},$$

Where $i = 1, ..., 389$ is an index of an acoustic feature, $n = 1, ..., 12$ is the index of the subject's answer to the interview question.

Let R be the matrix of acoustic characteristics values of two text fragments readings by the tested person:

$$R = \begin{pmatrix} r_{11} & ... & r_{1i} \\ r_{21} & ... & r_{2i} \end{pmatrix}.$$

T^1 is defined as the matrix of differences between the acoustic characteristics of the examinee's answers and acoustic characteristics of reading text #1:

$$T_n^1 = I_n - R_1.$$

T^2 is the matrix of differences between the acoustic characteristics of the examinee's answers and acoustic characteristics of reading text #2:

$$T_n^2 = I_n - R_2.$$

Then let M^1 be the matrix of averaged acoustic characteristics of the examinee's answers and his acoustic characteristics of reading text #1:

$$M_n^1 = \frac{I_n + R_1}{2}.$$

Then let M^2 be the matrix of averaged acoustic characteristics of the examinee's answers and his acoustic characteristics of reading text #2:

$$M_n^2 = \frac{I_n + R_2}{2}.$$

Table 1.

Results of model training by ROC-AUC score on the Full Data Set (Union Data)

Scale	Gaussian Process	Gradient Boosting Classifier	Linear SVM	K Nearest Neighbors	Poly SVM	QDA	Random Forest	RBF SVM
A	0.62	0.62	**0.68**	0.57	0.50	0.59	0.61	0.58
AM	0.56	**0.59**	0.56	0.55	0.50	0.56	0.55	0.50
AP	0.50	**0.67**	0.66	0.60	0.50	0.60	0.64	0.57
BM	0.50	**0.54**	0.53	0.51	0.50	0.52	0.50	0.51
BP	0.50	0.49	0.48	0.51	0.50	0.50	**0.53**	0.50
C	0.61	0.63	**0.67**	0.61	0.64	0.61	0.62	0.66
DM	0.63	0.643	**0.645**	0.57	0.50	0.57	0.61	0.63
DP	0.57	0.58	0.56	0.55	0.50	0.56	**0.59**	0.51
E	0.50	0.55	**0.58**	0.54	0.50	0.54	0.55	0.52
GM	0.49	0.56	0.46	**0.57**	0.50	0.55	0.53	0.50
GP	0.50	0.56	0.56	0.52	0.50	0.54	0.55	**0.57**
N	0.54	**0.58**	0.56	0.56	0.50	0.57	0.52	0.54
O	0.50	**0.53**	0.51	0.47	0.50	0.48	0.52	0.52
FIQ	0.50	**0.550**	0.549	0.50	0.50	0.52	0.54	0.50
CIQ	0.50	**0.55**	0.51	0.51	0.50	0.51	0.51	0.49

The number written in bold is the highest in the row.

Let K be the vector of averaged acoustic characteristics for all answers of the tested person:

$$K_i = \frac{\sum_{v=1}^n I_{vi}}{n}.$$

The corresponding matrices for all subjects were combined into training and test samples in which the subjects did not overlap.

B. Basic model selection

The choice of the basic model was made

among the following machine learning models: Gaussian process classification (GPC), Gradient Boosting Classifier, Linear SVM, K-Neighbors Classifier, Poly SVM, Quadratic Discriminant Analysis, Random Forest, RBF SVM for all psychodiagnostic techniques using a complete data set (*Union data*), including acoustic characteristics of text reading and audio-interview of the subjects. The division into training and test samples was carried out according to respondents, i.e. the training sample did not include recordings of the subject

who fell into the test sample. Thus, it was guaranteed that the model would not overfit on the data from a particular user.

The results of model training without parameter fine-tuning were analyzed by the ROC-AUC score and are presented in Table I.

The model based on the Gradient Boosting Classifier showed the best result for the majority of psychological characteristics, so we have chosen it as a basic one. The effect of different data preprocessing pipelines combined with the GBC method will be clarified further.

C. The psychological meaning of datasets

Different kinds of data processing (i.e., Union data, I, M^1, M^2, T^1, T^2) had a different meaning from the psychological point of view. Answering interview questions and reading Likhachev's memories about the siege of Leningrad presumed deeply emotional involvement. Reading an extract from science fiction was assumed to be emotionally neutral and could be relevant to the basic acoustic characteristics of an individual voice. Interview data represented voice features in personally significant situations. We proposed that such kind of self-presentation should be most relevant to the manifestation of personal traits.

Adding to Interview data any kind of reading data broadened the range of voice properties. The M^2 dataset extended the range of voice properties towards the emotional end, while the M^1 dataset increased the variability of neutral characteristics. The Union dataset incorporated the widest range of acoustic characteristics across different situations.

On the contrary, the T^1 and T^2 datasets (which were Text 1 and Text 2 feature matrices subtracted from the Interview feature matrix) restricted the range of vocal characteristics. We assumed that the acoustic characteristics of the voice in neutral intonation could have an individual profile (for example, neutral tonality is different for people with a strong manifestation of extraversion or introversion). Thus, the dataset T^1 made it possible to identify the acoustic characteristics that are most pronounced in the interview compared to the respondent's personal neutral tone. Similarly, the dataset T^2 allowed us to highlight the emotional manifestations in the speaker's profile compared to the reading of an emotionally loaded text.

The main psychological hypothesis underlying the selection of different data sets was that in order to diagnose personality traits, it was necessary to take into account the voice properties demonstrated by the respondent in different situations, highlighting the most significant deviations from the neutral tone. It was also necessary to consider the recording conditions (the task that the respondent is given), which could also influence the quality of psychological traits diagnostics. For example, if the respondent recorded the voice in a simulated dating situation, then it was more likely that the properties manifested in the voice will differ from those in a hiring situation, because the person would unconsciously try to demonstrate some of his or her features through the voice, introducing some distortion in the voice properties. We considered that recording the neutral text reading would prevent the respondent from introducing this distortion and thus provide a clear baseline for voice features.

Results

A. Personality traits classification results

A comparison of the classification quality measured by ROC-AUC score was conducted on all previously described datasets (T^1 and T^2 - the differences of the acoustic characteristics from reading texts 1 and 2 respectively; M^1 and M^2 - the averaged acoustic characteristics with reading texts 1 and 2 respectively; I - the initial acoustic characteristics without including data from text reading). The models based on the Gradient Boosting Classifier were trained separately for men and women, since it is assumed that men and women differ in the manifestation of psychological features through the acoustic parameters of speech.

Fig. 5 shows the model performance for women. The model trained using the answers to the interview questions (I) showed the best results for the following 5 scales out of 15: Conscientiousness (0.71), Extraversion (0.58), Beta-minus (0.51), Beta-plus (0.59), Delta-minus (0.697). A model trained using averaged acoustic characteristics with text 2 reading scores (Likhachev Memories) (M^2) showed the best results for the following 7 scales: Openness (0.56), Agreeableness (0.67), Neuroticism (0.64), Alpha-minus (0.67), Alpha-plus (0.70), Gamma-minus (0.64), Fluid Intelligence (0.55). The difference in acoustic features with text 1 Stanislav Lem "Solaris" (T^1) showed the best

results for Gamma-plus (0.64) and verbal intelligence (0.55) scales, and for text 2 "Likhachev Memories" (T^2) for Delta-plus scale (0.69). The numerical values of the ROC AUC scores are given in Appendix Table I.

Fig. 6 shows the model performance for men. The model trained using the answers to the interview questions (I) showed the best results for the following 2 scales out of 15: Alpha-minus (0.558), Delta-minus (0.716). The model trained using averaged acoustic characteristics with the reading scores of Stanislav Lem's text 1 "Solaris" (M^1) showed the best results for the Conscientiousness (0.759) and Gamma-plus (0.646) scales, and for text 2 (Likhachev Memories) (M^2) for the following scales: Openness (0.697), Agreeableness (0.588), Beta-plus (0.677), Gamma-minus (0.567). The difference in acoustic features for text 1 Stanislav Lem's "Solaris" (T^1) showed the best results for the scales Extraversion (0.566), Neuroticism (0.619), Alpha-plus (0.666), Crystallized Intelligence (0.701), and for text 2 "Likhachev Memories" (T^2) for Fluid Intelligence scale (0.576). The numerical values of the ROC AUC scores are given in Appendix Table II.
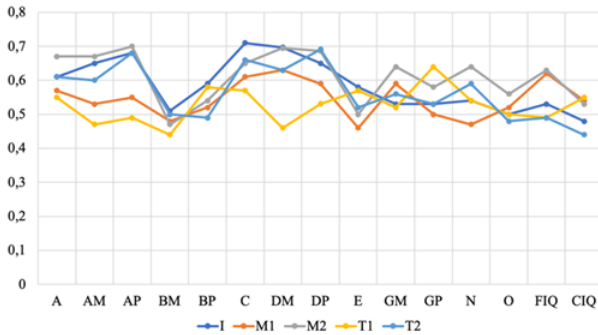
Figure 5. ROC-AUC score for Gradient Boosting Classifier - based model using different initial datasets for women.
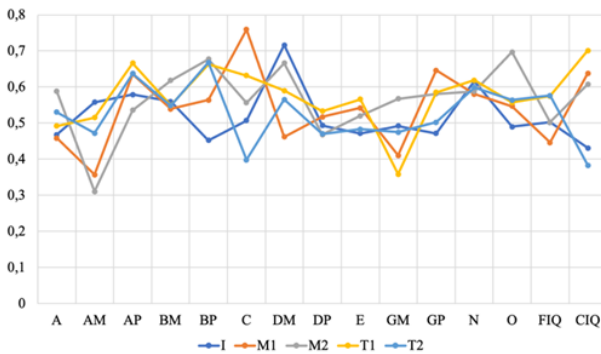
Figure 6. ROC-AUC score for Gradient Boosting Classifier - based model using different initial datasets for men.

The proposed approach for personality psychological traits diagnostics has been shown successful on a different set of personality traits depending on gender. Further comparison was made between the performance of the models for the original data (I) and all of the data modification options under consideration.

In models of psychological traits of *women*, an increase in prediction ROC-AUC of 10% (M^2) was observed for the neuroticism scale and Gamma-plus and Gamma-minus scales of 11% (T^1, M^2) related to mental health, subjective well-being. There was a 9% (M^2) increase in the ROC-AUC of nonverbal ability classification when using averages with reading text 2. Determining the ROC-AUC of the Agreeableness scale increased by 7% (M^2). In 7 of 15 cases, using averages with text reading (M^2) showed the best results for the classification of psychological traits.

In models of *men* personality traits, the greatest increase of 27% (T^1) was observed for the Nonverbal Intelligence scale. The Integrity scale showed an increase of 25% (M^1), and the Beta-plus and Beta-minus scales showed a 22% (M^2) and 10% (G) increase, respectively. The Openness to Experience scale showed an increase of 21% (M^2). Accuracy for the Gamma-plus and Gamma-minus scales increased 18% (M^1) and 8% (M^2), respectively. The Agreeableness and Extraversion scales showed increase of 12% (M^2) and 10% (T^1) respectively. These results do not allow us to identify a single preprocessing approach that would demonstrate the best classification results in most cases, however, an increase in classification accuracy was shown after applying different approaches.

B. Gradient Boosting vs other classification algorythms

As shown earlier in Table I, in some cases the Gradient Boosting Classifier did not show the best results in terms of ROC AUC score. Thus we trained a set of other algorithms using the winning models from Table I on different sets of preprocessed data. We compared the results obtained with the ones of Gradient Boosting Classifier (see Figs. 5 and 6 and Appendix Tables I and II) on different types of the preprocessed data.

Tables II and III show the results of different groups of models:

– #1 Models contain the best results for the Gradient Boosting Classifier on preprocessed data (Figs. 5 and 6; Appendix Tables I and II);

– #2 Models contain the most optimal algorithm from Table I, but applied to all types of the preprocessed data;

– #3 Models contain the training results of merged interview and reading data without preprocessing.

The use of #3 Model allowed us to analyze the effectiveness of the proposed approach to accounting for the audio recording scenario. This is, in case Model 3 won in diagnosing some psychological trait, we could conclude that the task and the recording circumstances did not matter in diagnosing this trait, i.e. our hypothesis was wrong. The results of Models #1 or #2 were interesting in terms of identifying the types of data preprocessing that showed the best result for each of the models.

Let's analyze the first line from Table II: according to Fig. 5 on the Agreeableness scale, the Gradient Boosting Classifier (#1 Models) showed a result of 0.67 (M²), in Table I we see the winning Linear SVM model for this trait. We further trained the models using all data sets using the Linear SVM algorithm (#2 Model), then compared their performance with the results obtained on the merged data set (#3 Model). The results of this comparison for *women* are shown in Table II.

Table II shows that for the Openness and Crystallized Intelligence traits specific audio processing and recording situation features did not matter, neither did text type of the task (reading or interview). The Beta-plus, Beta-minus, Conscientiousness, Delta-minus, Extraversion traits were not reflected in the acoustic properties of speech, which differed in text reading and in self-presentation task. However, it is important that it was possible to diagnose the features in the interviewing process, but not in the reading process. The scales Agreeableness, Alpha-plus, Alpha-minus, Gamma-minus, Neuroticism, Openness showed the best results when using averaged acoustic characteristics with the text reading 2 "Likhachev Memories" (M²), which confirms the proposed hypothesis. Delta-plus and Gamma-plus traits demonstrated the best results when using the data sets T² and T¹, respectively, which indicates the significance of accounting for the difference in acoustic characteristics between interviewing and reading.

A similar comparison of models trained using different algorithms for *men is* shown in Table III.

According to Table III, it can be noted that for Alpha-minus and Fluid intelligence traits the hypothesis tested was not confirmed: the scenario of interview recording or text reading did not play a significant role for the diagnostic model. The diagnostic models for the Agreeableness and Delta-minus traits showed the best performance during the interview, without taking into account text reading characteristics. We see that in *men* the highest ROC-AUC scores were observed for the data on the average acoustic characteristics of the audio-interview and text reading 1 - Stanislav Lem "Solaris" (M¹) for the Conscientiousness, Extraversion, Gamma-minus, Gamma-plus traits. Average acoustic characteristics of the audio-interview and text reading 2 - "Likhachev's Memories" (M²) showed the best results for Beta-plus and Openness scales. Alpha-plus, Delta-plus, Crystallized Intelligence traits showed the best results for the dataset T¹ (the difference between the values of acoustic characteristics of audio-interview and reading text 1 Stanislav Lem "Solaris").

C. Male – female differences

When comparing the accuracy results of the models for men and women, it should be noted that the Conscientiousness and Delta-minus traits were best diagnosed by analyzing the acoustic characteristics of audio-interview recordings, as well as by the M¹ dataset for both men and women with similar ROC-AUC values.

In females, the ROC-AUC threshold = 0.7 was achieved for the Alpha-plus trait on the M² dataset, in males it was achieved for Openness (M²) and Crystallized Intelligence (T¹).

It should be noted that the Gamma-plus and Gamma-minus traits were represented among women on the winning models in the datasets (T¹ or M¹), associated with reading the text 1 - Stanislav Lem "Solaris". For men, similarly, the data set M¹ (text 1 - Stanislav Lem "Solaris") proved to be the best for these two traits, and for the traits Extraversion, Delta-plus, Alpha-plus, Crystallized Intelligence, Neuroticism the consideration of characteristics in reading text 1 demonstrated improved results.

Table 2.

Classification results. The table reports the ROC-AUC scores for different kinds of algorithms and input data for women

Scale	#1 Models using preprocessing data			#2 Models with algorithm-winner using preprocessing data			#3 Models using original union data		
	ROC-AUC	Data	Model	ROC-AUC	Data	Model	ROC-AUC	Data	Model
A	**0.67**	M[2]	Gradient Boosting	0.66	M[2]	Linear SVM	0.64	Interview& Reading	Gradient Boosting
AM	**0.67**	M[2]	Gradient Boosting	-	-	-	0.62	Interview& Reading	Gradient Boosting
AP	**0.70**	M[2]	Gradient Boosting	-	-	-	0.64	Interview& Reading	Gradient Boosting
BP	**0.59**	I	Gradient Boosting	0.57	T[1]	Random Forest	0.51	Interview& Reading	Gradient Boosting
BM	**0.51**	I	Gradient Boosting	-	-	-	0.47	Interview& Reading	Gradient Boosting
C	0.71	I	Gradient Boosting	**0.75**	I	Linear SVM	0.64	Interview& Reading	Gradient Boosting
DM	0.69	I	Gradient Boosting	**0.72**	I	Linear SVM	0.63	Interview& Reading	Gradient Boosting
DP	**0.69**	T[2]	Gradient Boosting	0.64	M[2]	Random Forest	0.61	Interview& Reading	Gradient Boosting
E	0.58	I	Gradient Boosting	**0.66**	I	Linear SVM	0.57	Interview& Reading	Gradient Boosting
GM	0.63	M[2]	Gradient Boosting	**0.64**	M[1]	K Nearest Neighbors	0.55	Interview& Reading	Gradient Boosting
GP	**0.64**	T[1]	Gradient Boosting	0.57	T[1]	RBF SVM	0.58	Interview& Reading	Gradient Boosting
N	**0.64**	M[2]	Gradient Boosting	-	-	-	0.51	Interview& Reading	Gradient Boosting
O	0.56	M[2]	Gradient Boosting	-	-	-	**0.58**	Interview& Reading	Gradient Boosting
FIQ	**0.63**	M[2]	Gradient Boosting	-	-	-	0.57	Interview& Reading	Gradient Boosting
CIQ	0.550	T[1]	Gradient Boosting	-	-	-	**0.552**	Interview& Reading	Gradient Boosting

The number written in bold is the highest in the row

Discussion and conclusions

In our study, we have shown the possibility to predict psychological characteristics (personality traits and intelligence) by audio analysis of voice characteristics. Unlike previous studies, we took into account not only personality traits according to the Big Five model, but also considered another personality model (The Circumplex of Personality Metatraits) as well as crystallized and fluid intelligence.

Our data suggest that different processes underline the manifestation of personal traits in voice properties in males and females. Not only the same traits were predicted with different accuracy, but the same datasets provided different accuracy (at least for personality traits). For females, the most relevant data for personality traits diagnosis were those involving different emotional contexts (self-presentation and empathy). For males, it was important to include a variety of situations to broaden the context as much as possible.

We found that in general, the identification of personality traits by acoustic characteristics of speech was more effective than the identification of intelligence. However, the prediction of intelligence was more consistent across men and women. Thus, we should search for more relevant context for intelligence diagnosis. From our data, it is evident that situations concerning self-presentation are less appropriate, but the text reading data contributes to enhanced classification quality. It is probable that more relevant data to predict intelligence could be obtained from think-aloud protocols or by using argumentation procedures.

It is also worth noting that the machine learning models considered were trained without parameter fine-tuning, which is one of the limitations of the current study and the direction of our future work.

It is also worth noting that the machine learning models considered were trained without parameter fine-tuning, which is one of the limitations of the current study and the direction of our future work.

Table 3.

Classification results. The table reports the ROC-AUC scores for different kinds of algorithms and input data for men

Scale	#1 Models using preprocessing data			#2 Models with algorithm-winner using preprocessing data			#3 Models using original union data		
	ROC-AUC	Data	Model	ROC-AUC	Data	Model	ROC-AUC	Data	Model
A	0.59	M^2	Gradient Boosting	**0.61**	I	Linear SVM	0.41	Interview& Reading	Gradient Boosting
AM	0.56	I	Gradient Boosting	-	-	-	**0.66**	Interview& Reading	Gradient Boosting
AP	**0.67**	T^1	Gradient Boosting	-	-	-	0.64	Interview& Reading	Gradient Boosting
BP	**0.68**	M^2	Gradient Boosting	0.59	M^2	Random Forest	0.62	Interview& Reading	Gradient Boosting
BM	**0.62**	M^2	Gradient Boosting	-	-	-	0.53	Interview& Reading	Gradient Boosting
C	**0.76**	M^1	Gradient Boosting	0.66	M^1	Linear SVM	0.64	Interview& Reading	Gradient Boosting
DM	**0.72**	I	Gradient Boosting	0.68	I	Linear SVM	0.69	Interview& Reading	Gradient Boosting
DP	0.53	T^1	Gradient Boosting	**0.56**	T^1	Random Forest	0.55	Interview& Reading	Gradient Boosting
E	0.57	T^1	Gradient Boosting	**0.59**	M^1	Linear SVM	0.52	Interview& Reading	Gradient Boosting
GM	0.57	M^2	Gradient Boosting	**0.67**	M^1	k Nearest	0.43	Interview& Reading	Gradient Boosting
GP	**0.65**	M^1	Gradient Boosting	0.61	M^2	RBF SVM	0.54	Interview& Reading	Gradient Boosting
N	**0.62**	T^1	Gradient Boosting	-	-	-	0.6	Interview& Reading	Gradient Boosting
O	**0.70**	M^2	Gradient Boosting	-	-	-	0.62	Interview& Reading	Gradient Boosting
FIQ	0.576	T^2	Gradient Boosting	-	-	-	**0.582**	Interview& Reading	Gradient Boosting
CIQ	**0.70**	T^1	Gradient Boosting	-	-	-	0.66	Interview& Reading	Gradient Boosting

The results show that the proposed approach to increasing the estimation accuracy of psychological traits of the person (using the audio-characteristics obtained in different types of tasks) sometimes appears to be more effective than the use of self-presentation audio-recordings only. We attribute this to the fact that certain properties are expressed by the person in the process of social interaction only and are related to the context of a situation.

It is of interest that the determination of the high and low level of expression of the psychological trait Conscientiousness showed the best quality among others for both men (M^1 dataset) and women (I dataset). This may be related to the simulated employment situation, where people try to demonstrate Conscientiousness to a greater extent among other personality traits. Also for both men and women, the best results on specific datasets related to reading Stanislav Lem's Solaris (M^1, T^1 datasets), which had a neutral tone, were obtained for the Gamma-plus and Gamma-minus scales, which characterize prosocial orientation, mental health, and self-control. The diagnosis of a number of traits in women was more extensively connected with the peculiarities of reading the emotionally loaded text "Likhachev's Memoirs," which was probably associated with a greater internal response to the tragic situation described in the text. For men, in turn, it was the reading of a neutral fragment of a science fiction novel that allowed to increase the accuracy of the diagnosis of a number of psychological traits.

In order to interpret the results obtained, it is necessary to determine the theoretical model underlying the relationships between voice properties and personality characteristics found in literature. However, almost all works focus exclusively on the studying of features. Mallory and Miller [3] have suggested that voice features (closely related to muscle reactions) and their corresponding personality traits develop in parallel as a result of a set of reactions to certain life events. For example, the situations of submission, falling into which leads to appropriate personality trait development, may

be accompanied by compression of the muscles regulating the vocal cords, leading to the formation of a higher voice, while narrowing of the vocal passages leads to a decrease in resonance properties. Further research has shown that the characteristics of the speech signal depend on the autonomic and somatic nervous system, and the vagus nerve, which supports motor parasympathetic fibers and is responsible for controlling heart rate and sweating, controls the activation of some muscles of the mouth and larynx [14].

Another approach, proposed by Silnitskaya [29], postulates that the connection between psychological characteristics and voice is moderated by the context of a person's activity. She showed that some temperamental and personal characteristics correlate differently with voice features depending on communication context (with or without interlocutor). In our study we found further evidence for Silnitskaya's theory, showing that the relation between voice and psychological characteristics is moderated not only by gender, but also by context in which speaking activity takes place.

Thus, we can conclude that further development of approaches to psychodiagnosis of personality traits through the analysis of subjects' speech should be related to the context and situation of recording of the subject's voice, as this may have a significant impact on the quality of models developed with a more complex structure.

Appendix

Table 1.

Classification Results. The Table Reports the ROC-AUC Scores for Different Kinds of Input Data for Females

Scale	I	M^1	M^2	T^1	T^2	Union
A	0,61	0,57	**0,67**	0,55	0,61	0.64
AM	0,65	0,53	**0,67**	0,47	0,60	0.62
AP	0,68	0,55	**0,70**	0,49	0,68	0.64
BM	**0,51**	0,48	0,47	0,44	0,50	0.51
BP	**0,59**	0,52	0,54	0,58	0,49	0.47
C	**0,71**	0,61	0,65	0,57	0,66	0.64
DM	**0,697**	0,630	0,695	0,46	0,63	0.63
DP	0,65	0,59	0,686	0,531	**0,691**	0.61
E	**0,58**	0,46	0,50	0,57	0,52	0.57
GM	0,53	0,59	**0,64**	0,52	0,56	0.55
GP	0,53	0,50	0,58	**0,64**	0,53	0.58
N	0,54	0,47	**0,64**	0,54	0,59	0.51
O	0,50	0,52	**0,56**	0,50	0,48	**0.58**
FIQ	0,53	0,62	**0,63**	0,49	0,49	0.57
CIQ	0,48	0,54	0,53	**0,55**	0,44	**0.552**

The number written in bold is the highest in the row.

Table 2.

Classification Results. The Table Reports the ROC-AUC Scores for Different Kinds of Input Data for Males

Scale	All	M^1	M^2	T^1	T^2	Union
A	0,468	0,458	**0,588**	0,492	0,530	0.41
AM	**0,558**	0,357	0,310	0,515	0,472	0.66
AP	0,579	0,636	0,536	**0,666**	0,637	0.64
BM	0,560	0,539	**0,618**	0,549	0,548	0.62
BP	0,452	0,564	**0,677**	0,663	0,668	0.53
C	0,507	**0,759**	0,556	0,632	0,398	0.64
DM	**0,716**	0,462	0,666	0,590	0,565	0.69
DP	0,493	0,517	0,468	**0,533**	0,470	0.55
E	0,471	0,542	0,520	**0,566**	0,483	0.52
GM	0,492	0,410	**0,567**	0,358	0,475	0.43
GP	0,471	**0,646**	0,581	0,585	0,502	0.54
N	0,616	0,580	0,588	**0,619**	0,599	0.6
O	0,490	0,547	**0,697**	0,557	0,564	0.62
FIQ	0,502	0,446	0,502	0,575	**0,576**	0.582
CIQ	0,431	0,638	0,608	**0,701**	0,383	0.66

The number written in bold is the highest in the row.

References

1. Allport, G. W., & Cantril, H. (1934). Judging personality from voice. *J. Social Psychology, 5*(1), 37–55. https://doi.org/10.1080/00224545.1934.9921582

2. An, G., & Levitan, R. (2018) Lexical and acoustic deep learning model for personality recognition. In *Proc. Interspeech 2018* (pp. 1761–1765), Hyderabad, India.

3. Baturin, N. A., & Kurganskii, N. A. (2005). Creation and standardization of the intelligence test for middle school age. *Psikhologicheskaya Nauka Obrazovanie (Psychological Sci. Educ.), 10*(3), 74–85. (in Russian)

Baturin N.A., Kurganskij N.A. Razrabotka i standartizaciya testa intellekta dlya srednego shkol'nogo vozrasta // Psihologicheskaya nauka i obrazovanie. 2005. Tom 10. № 3. S. 74–85.

4. Biel, J.-I., & Gatica-Perez, D. (2013). The YouTube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs. *IEEE Trans. Multimedia, 15*(1), 41–55. https://doi.org/10.1109/tmm.2012.2225032

5. Block, J. (2010). The five-factor framing of personality and beyond: Some ruminations. *Psychological Inquiry, 21*(1), 2–25. https://doi.org/10.1080/10478401003596626

6. Bors, D. A., & Stokes, T. L. (1998). Raven's advanced progressive matrices: Norms for first-year university students and the development of a short form. *Educational Psychological Meas., 58*(3), 382–398. https://doi.org/10.1177/0013164498058003002

7. Carbonneau, M. A., Granger, E., Attabi, Y. & Gagnon, G. (2020). Feature Learning from Spectrograms for Assessment of Personality Traits. *IEEE Transactions on Affective Computing, 11*(1), 25–31. https://doi.org/10.1109/TAFFC.2017.2763132

8. Costa, P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment, 64*, 21-50.

9. Digman, J. M. (1997). Higher-order factors of the big five. *J. Personality Social Psychology, 73*(6), 1246–1256. https://doi.org/10.1037/0022-3514.73.6.1246

10. Feldstein, S. & Sloan, B. (1984). Actual and stereotyped speech tempos of extraverts and introverts. *J. Personality, 52*(2), 188–204. https://doi.org/10.1111/j.1467-6494.1984.tb00352.x

11. Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4*(1), 26–42. https://doi.org/10.1037/1040-3590.4.1.26

12. Guidi, A., Gentili, C., Scilingo, E. P., & Vanello, N. (2019). Analysis of speech features and personality traits. *Biomed. Signal Process. Control, 51*, 1–7. https://doi.org/10.1016/j.bspc.2019.01.027

13. John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy: History, measurement, and conceptual issues. In *Handbook of Personality: Theory and Research* (pp. 114–158).

14. Kreiman, J., & Sidtis, D. (2011). *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception.* Hoboken, NJ: John Wiley & Sons.

15. Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *J. Artificial Intell. Res., 30*, 457–500. https://doi.org/10.1613/jair.2349

16. Mallory, E. B., & Miller, V. R. (1958). A possible basis for the association of voice characteristics and personality traits. *Speech Monographs, 25*(4), 255–260. https://doi.org/10.1080/03637755809375240

17. Mauchand, M., & Pell, M. D. (2021). Emotivity in the voice: Prosodic, lexical, and cultural appraisal of complaining speech. *Frontiers Psychology, 11*, Article 619222. https://doi.org/10.3389/fpsyg.2020.619222

18. McFee, B., *et al.*, "librosa: Audio and music signal analysis in python. In *Proc. 14th Python Sci. Conf.* (pp. 18–25), Texas, USA.

19. Mehta, Y., Majumder, N., Gelbukh, A., & Cambria, E. (2019). Recent trends in deep learning based personality detection. *Artificial Intell. Rev., 53*(4), 2313–2339. https://doi.org/10.1007/s10462-019-09770-z

20. Mohammadi, G., Vinciarelli, A., & Mortillaro, M. (2010). The voice of personality: Mapping nonverbal vocal behavior into trait attributions. In *SSPW '10 - Proc. 2010 ACM Social Signal Process. Workshop Co-Located ACM Multimedia 2010* (pp. 17–20), Italy.

21. Panfilova, A., & Pospelov, N. (2022). A

reading and self-presentation speech characteristics dataset. *IEEE Dataport.* https://doi.org/10.21227/hrkm-wt26

22. Park, J., Lee, S., Brotherton, K., Um, D., & Park, J. (2020). Identification of speech characteristics to distinguish human personality of introversive and extroversive male groups. *Int. J. Environmental Res. Public Health, 17*(6), Article 2125. https://doi.org/10.3390/ijerph17062125

23. Polzehl, T. (2015). *Personality in Speech.* Springer.

24. Polzehl, T., Moller, S., & Metze, F. (2010). Automatically assessing personality from speech. In *2010 IEEE Fourth Int. Conf. Semantic Comput.* (pp. 134–140), Pittsburgh.

25. Ramsay, R. W. (1968). Speech patterns and personality. *Language Speech, 11*(1), 54–63. https://doi.org/10.1177/002383096801100108

26. Sapir, E. (1927). Speech as a personality trait. *American J. Sociology, 32*(6), 892–905. https://doi.org/10.1086/214279

27. Schuller, B., Steidl, S., & Batliner, A. (2009). The INTERSPEECH 2009 emotion challenge. In *Proc. Interspeech 2009* (pp. 312–315), Brighton, United Kingdom.

28. Shchebetenko, S. A. (2014). The best man in the world: Attitudes toward personality traits. *Psychology J. Higher School Economics, 11*(3), 129–148. (in Russian)

Shchebetenko S.A. «Luchshij chelovek v mire»: ustanovki na cherty lichnosti // Psihologiya. ZHurnal Vysshej shkoly Ekonomiki 2014, T.11, № 3, S.129–148

29. Silnitskaya, A. S., & Gusev, A. N. (2013). Character and temperamental determinants of prosodic parameters of natural speech. *Psychology in Russia*, 6(3), 95–106.

30. Stern, J., *et al.* (2021). Do voices carry valid information about a speaker's personality? *J. Res. Personality, 92,* Article 104092. https://doi.org/10.1016/j.jrp.2021.104092

31. Strus, W., & Cieciuch, J. (2017). Towards a synthesis of personality, temperament, motivation, emotion and mental health models within the Circumplex of Personality Metatraits.

J. Res. Personality, 66, 70–95. https://doi.org/10.1016/j.jrp.2016.12.002

32. Tatarko, A., Maklasova, E., & Grigoryan, K. (2019). Validation of the circumplex of personality metatraits questionnaire on the Russian sample. *Psychology J. Higher School Economics, 16*(4), 705–729 (in Russian) DOI: 10.17323/1813-8918-2019-4-705-729

Tatarko A.N., Maklasova E.V., Grigoryan K.A. Validizaciya oprosnika Krugovaya struktura lichnostnyh metachert na rossijskoj vyborke // Psihologiya. ZHurnal Vysshej SHkoly Ekonomiki 2019, T. 16, № 4, S. 705–729

33. Tayarani, M., Esposito, A. & Vinciarelli, A. (2019). What an "Ehm" leaks about you: Mapping fillers into personality traits with quantum evolutionary feature selection algorithms. *IEEE Trans. Affective Comput., 13*, 108–121. https://doi.org/10.1109/taffc.2019.2930695

34. Truesdale, D. M., & Pell, M. D. (2018). The sound of passion and indifference. *Speech Commun., 99,* 124–134. https://doi.org/10.1016/j.specom.2018.03.007

35. Vallabha, G. K. & Tuller, B. (2002). Systematic errors in the formant analysis of steady-state vowels. *Speech Commun., 38*(1-2), 141–160. https://doi.org/10.1016/s0167-6393(01)00049-8

36. Valueva, E. A., & Ushakov, D. V. (2010). Empirical verification of the model of relation of cognitive and emotional abilities. Psychology. *J. Higher School Economics, 7*(2), 103–114. (in Russian)

Valueva E.A., Ushakov D.V. Empiricheskaya verifikaciya modeli sootnosheniya predmetnyh i emocial'nyh sposobnostej // Psihologiya. ZHurnal vysshej SHkoly Ekonomiki 2010, T. 7, № 2, S. 103–114

37. Zvarevashe, K., & Olugbara, O. (2020). Ensemble learning of hybrid acoustic features for speech emotion recognition. *Algorithms, 13*(3), Article 70. https://doi.org/10.3390/a13030070

Author contributions. N. Pospelov was engaged in the extraction of the audio characteristics, A. Panfilova and D. Parkhomenko developed models, E. Valueva selected psychodiagnostic tools.

Competing interests. None