www.nature.com/scientificreports

scientific reports

Check for updates

OPEN

Speech-based personality prediction using deep learning with acoustic and linguistic embeddings

Martin Lukac

This study introduces a novel method for predicting the Big Five personality traits through the analysis of speech samples, advancing the field of computational personality assessment. We collected data from 2045 participants who completed a self-reported Big Five personality questionnaire and provided free-form speech samples by introducing themselves without constraints on content. Using pre-trained convolutional neural networks and transformer-based models, we extracted embeddings representing both acoustic features (e.g., tone, pitch, rhythm) and linguistic content from the speech samples. These embeddings were combined and input into gradient boosted tree models to predict personality traits. Our results indicate that personality traits can be effectively predicted from speech, with correlation coefficients between predicted scores and self-reported scores ranging from 0.26 (extraversion) to 0.39 (neuroticism), and from 0.39 to 0.60 for disattenuated correlations. Intraclass correlations show moderate to high consistency in our model's predictions. This approach captures the subtle ways in which personality traits are expressed through both how people speak and what they say. Our findings underscore the potential of voice-based assessments as a complementary tool in psychological research, providing new insights into the connection between speech and personality.

Keywords Personality, Speech, Machine learning, Psychological assessment, Computational social science

> Speech being delicate, subtle and powerful form of behaviour, the way in which a thing is said is often as important as the message. Pear, 1931.

In human communication, speech uniquely encapsulates not only conveyed messages but also facets of a speaker's identity, from age and cultural nuances to deeper layers of personality[1]. Historically, voice has intrigued as a gateway to personality, with discernment attempts tracing back to the 1930s[2]. Today, the synergy of advanced technology, data access, and artificial intelligence amplifies this exploration, ushering in opportunities to predict personality traits from voice. Previous research has established that voice carries information about one's psychological process through pitch, volume, timbre, or tone of the sound produced[3–5]. Moreover, in combination with the content of the speech, samples of human speech have the potential to become a valid and reliable source of personality predictions.Building on this historical context, computational personality assessments have gained traction in recent decades as novel and reliable means of assessing personality. Ranging from "digital footprints" derived from social networks (e.g. Facebook or Twitter)[6] to personal communication (e.g. WhatsApp), including static facial images[7], eye movements[8,9], or text in general[10], computational personality assessment is becoming an established endeavour. Such research has enabled subsequent advances showing how predicted personality can help match people to their ideal jobs[11], or how founder personalities impact startup success[12].

While significant associations between various inputs and personality traits have been identified, their predictive strength is often limited or the necessary data is not readily accessible at the time of assessment. For instance, obtaining a job applicant's social media activity, such as Facebook or Twitter posts, for analysis is not feasible due to privacy concerns and ethical considerations. Similarly, using eye-tracking data for personality assessment requires specialized equipment and controlled environments, making it impractical in typical business settings. These limitations, along with ongoing debates over the validity and ethical implications of such methods, pose challenges for their practical application, particularly in business contexts where they could significantly impact decision-making[13].

In this paper, we evaluate whether individuals' Big Five trait levels can be predicted on the basis of speech samples. Personality, in this case, refers to the enduring patterns of thoughts, feelings, and behaviours that distinguish individuals from one another[14]. In our study, we focus on two primary aspects of speech: the

Koios, London, UK. email: martin@getkoios.ai

acoustic features, which include elements like pitch and tone, and the linguistic features, which involve the actual content of the speech and examine their mutual ability to predict personality traits. We assessed personality in terms of the Big Five personality traits—the most widely used, validated, and well-established personality trait theory in psychological science. The theory describes human personality in terms of five factors: agreeableness, conscientiousness, extraversion, neuroticism (often referred to as emotional stability), and openness (or openness to experience). The Big Five personality theory is the most robust and parsimonious model to understand personality[15]. It has been widely validated and replicates across various demographics and cultures[16,17], and research shows Big Five traits remain relatively stable in time[18–20].

There are several theoretical reasons for speech samples to carry valid signals about speakers' personality (further elaborated in[4]). Personalities influence individual's expressions: emotional, intellectual, and even physical engagements shape conversational content. Previous studies found systematic empirical associations between personality and individual differences in word use[21,22]. Furthermore, the nuances of speech delivery—like intonation, rhythm, and volume—are subtly moulded by an individual's personality[23,24]. An extraverted person, for example, might display a more dynamic and resonant vocal pattern, a direct projection of their sociable nature. Certain external variables, like cultural background, education, or upbringing, could concurrently shape both an individual's personality and their speech patterns.

Recent studies have explored these relationships by examining verbal and paraverbal behaviours in relation to personality within interview settings. Koutsoumpis and de Vries found that voice characteristics are informative for traits such as openness to experience, emotionality, and conscientiousness, highlighting that traits are activated in relevant contexts and expressed through changes in voice[25]. Similarly, Koutsoumpis et al. recently utilised machine learning techniques to analyse asynchronous video interviews, automatically extracting words, facial expressions, and voice characteristics to predict personality traits and interview performance[26]. Their findings demonstrate that combining verbal and non-verbal cues enhances the accuracy of personality assessments, aligning closely with the aims of our current study and suggesting that leveraging modern computational methods can improve the assessment of personality from speech, particularly in contexts that mimic real-world interactions.

A variety of demographic and cultural factors might influence both an individual's personality and their vocal characteristics. For instance, sex differences are known to affect both voice—such as pitch and tone—and personality traits, with research indicating reasonably large differences at the facet level[27,28]. Age is another factor that impacts voice qualities due to physiological changes over time and is also associated with shifts in personality traits[29]. Additionally, language proficiency, such as speaking English as a second language, and related ethnic, racial, or cultural backgrounds, can influence speech patterns[30]. These variables can simultaneously shape how individuals express themselves vocally and their personality traits, potentially confounding the direct relationship between voice and personality.

Self-reported personality assessments, though widely used, have well-documented biases and limitations—such as social desirability bias[31], response styles[32], lack of self-awareness[33], and intentional distortion (faking)[34]—that can compromise their validity. Voice-based assessments may help mitigate some of these issues by analysing behavioural expressions of personality through speech patterns, potentially capturing objective acoustic and linguistic features that are less susceptible to conscious manipulation[35]. While individuals might attempt to adjust aspects of their speech, such as volume or content, especially in high-stakes settings, many vocal characteristics (e.g., subtle intonations, pitch variability, speech rhythm) are difficult to control intentionally and may reveal underlying personality traits[36]. Furthermore, although our model is trained on self-reported personality scores—which may themselves contain social desirability biases—the aim is to develop an alternative assessment method that leverages observable behaviour to predict personality traits. This approach may not eliminate the influence of biases but offers a tool less reliant on self-reporting and more grounded in behavioural indicators, thereby helping to complement self-reports by providing additional perspectives on personality assessment.

Historically, the assessment of personality from speech has largely been determined through human evaluations. Research in this domain traces back to the 1930s when listeners judged radio voices for their personalities[2]. Most studies, with a few exceptions (e.g.[36]), have relied on human judges to infer personality traits from speech. With advancements in technology, however, contemporary research has begun investigating the correlation between specific vocal parameters and personality traits[37]. While standardised voice recordings offer consistent parameters, they negate the role of language by compelling all participants to use identical words. On the other hand, naturalistic voice samples, although more representative, present challenges due to variability in contexts[35] and are harder to gather on a larger scale. A limitation shared across the literature attempting to map acoustic characteristics to personality traits is the manual extraction of features from speech samples of predefined length (see[4,5,35,37]). Novel neural network architectures can potentially overcome these limitations, but are known to require large labelled training datasets that are difficult and costly to collect.

The purpose of this study is to introduce a new approach that tackles the existing limits of computational personality trait prediction from speech samples. We address the issues laid out above by collecting a novel dataset ($N = 2045$) of personality assessments and samples of free speech of individuals from a survey pool provided by Prolific, showing one of the highest data quality across behavioural data providers online[38]. Our sample is representative of the UK population in age, gender, ethnicity, and first language distributions, ensuring fair representation. Respondents were informed about the purpose of the data collection and were guaranteed to be paid higher than minimum wage for their responses. Each respondent was asked to answer 50 questions from the International Personality Item Pool (known as IPIP), 10 questions per personality trait. After that, we asked respondents to introduce themselves in free speech, completely unconstrained in what they decide to speak about.

In this study, we analyse speech through two primary lenses: acoustic and linguistic features. Acoustic features were extracted using the embedding vector of a pre-trained convolutional neural network (CNN), capturing aspects such as pitch, tone, and rhythm, which are expected to remain consistent across different speech contexts. Linguistic features were derived from the textual content of the speech, using the embedding of the encoder of a large language model (LLM), and are recognised as being more dependent on the nature of the prompts given and the specific responses provided by the participants. These embedding vectors are then concatenated and funnelled into a second-layer of models which are trained to predict individual traits (one model per trait).

In the first-layer models, focused on extracting embeddings, the scarcity of the data can be ameliorated by not training models from scratch, but using transfer learning and by using models pretrained on auxiliary audio classification tasks[39,40]. Examples of using pretrained models for audio classification include Parkinson's disease early detection[41], predicting vocal reactions to music[42], or wildlife protection and preservation[43–45]. Similarly for text, approaches using text embeddings extracted from pretrained and open-sourced models (e.g. word2vec) showed improved accuracy over models trained from scratch[46]. Modern pretrained transformer architectures with attention mechanisms, commonly available in open-source, immensely improved text classification tasks for cases where data are hard to obtain (see[47,48]). The second-layer models in our architecture are gradient boosted trees, trained on the collected data to map the extracted acoustic and linguistic embeddings to individual personality traits (see Fig. 1 for a schematic summary). Hyperparameters of these models were tuned using fivefold cross-validation using only the training set, and then subsequently evaluated on a hold-out test set. This methodological framework, integrating convolutional neural networks, transformer-based architectures, and gradient boosted trees for personality trait prediction, sets the stage for the ensuing results section, which presents the empirical findings from our dataset.

Results

Prediction accuracy

Previous meta-analytic studies suggest that computational personality assessment performances vary widely, often influenced by the nature of the data and the complexity of the models employed. Evaluations of model performance depend on the specific goals, such as whether the aim is to replace traditional personality measures or to complement them[49]. In the case of replacing traditional measures, models should ideally match or surpass the self-other correlations (correlation between the *self*-assessment and personality assessment of a target-person by *other* people; $\rho = 0.29–0.41$)[50] or between-inventories correlations ($\rho = 0.31–0.56$)[51].

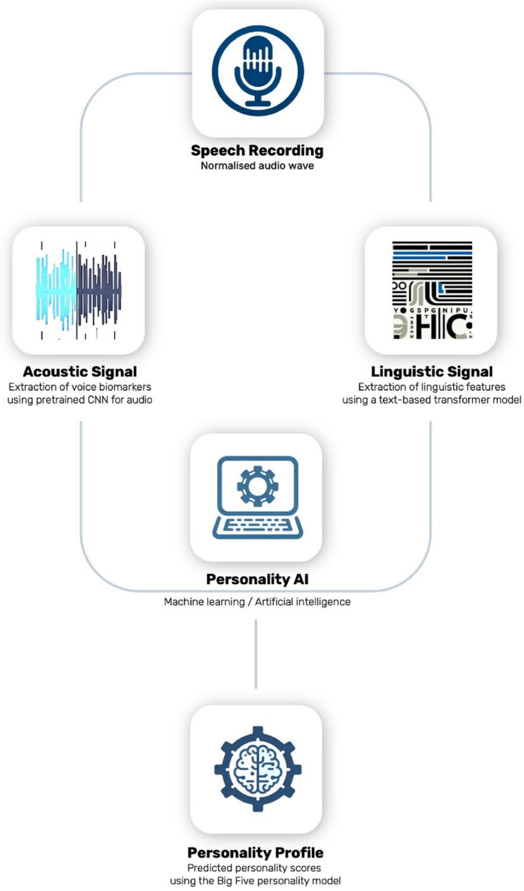

Figure 1. A schematic of the proposed model to predict the Big Five personality profile from voice, using acoustic and linguistic embedding features.

The correlation coefficients between the self-reported test scores and the scores predicted by the model proposed in this paper ranged from 0.26 to 0.39. The associations were strongest for neuroticism (0.39) and agreeableness (0.38) and lowest for extraversion (0.26) (see Fig. 3). The observed correlations between model-predicted scores and self-reported test scores for neuroticism and agreeableness align with the theoretical underpinnings discussed earlier, suggesting that certain personality traits are more readily expressed and discernible in speech patterns. The strong association for neuroticism could be attributed to its inherent link with emotional expressivity, often reflected in vocal nuances such as pitch variability and speech rate. Agreeableness, characterised by cooperative and compassionate behaviour, might similarly manifest in speech through tone and language use. The lower correlation for extraversion (0.26), however, presents an interesting divergence from expectations. Given the societal perception of extraversion as outwardly observable through social interactions, its subtler nuances in speech may not be as easily captured by the current models. This discrepancy highlights the complexity of vocal expression as a mirror of personality, influenced by factors including cultural background, personal experiences, and the context of the speech itself.

Upon closer examination of the demographic influences on our model's predictive accuracy (see Supplementary Material), the results indicate a negligible difference in accuracy across genders for all personality traits. Both males and females exhibit comparable correlation coefficients, suggesting an equitable predictive performance by the model for both genders. Regarding language bias, the differences in prediction accuracy between English and non-English speakers as their first language are insignificant for openness, neuroticism, and agreeableness. However, there is an observed difference in accuracy for conscientiousness and extraversion, where the predictive accuracy is lower for speaker whose first language is not English. This is likely caused by the relatively lower number of non-English speakers in the sample. This indicates the need to up-sample this demographic in the training sample to achieve fully equitable performance across different speaker groups.

In light of the mentioned criteria, the present study's findings, demonstrating correlation coefficients between 0.26 and 0.39 suggest a promising alignment with the aforementioned benchmarks.

Consistency and reliability

To further evaluate the consistency and reliability of our model's predictions, Intraclass Correlations (ICC) were calculated. ICC is a statistical measure used to assess the degree of agreement or conformity among ratings made by different raters or measurements. In our study, the ICC specifically measures how closely the predicted personality trait scores from one half of a participant's recording align with the predictions from the other half of the same recording (more details in "Calculation of consistency and reliability"). The ICC values for the Big Five traits indicated a moderate to high level of consistency, with agreeableness at 0.725, conscientiousness at 0.591, extraversion at 0.645, neuroticism at 0.713, and openness at 0.573 (shown in Fig. 2).

To provide a direct comparison with the reliability of the self-report measures, we calculated the split-half reliability coefficients of the IPIP Big Five scales used in our study, applying the Spearman-Brown prophecy formula. The split-half reliability coefficients were high across all traits, with agreeableness at 0.886, conscientiousness at 0.857, extraversion at 0.925, neuroticism at 0.927, and openness at 0.815. These values indicate excellent internal consistency of the self-report instruments. Comparing these results, we observe that the ICC values of the proposed models are lower than the split-half reliability coefficients of the self-report scales. This difference is expected, given the inherent variability in speech patterns. Nevertheless, the moderate to high ICC values demonstrate that the models produce reliable predictions of personality traits from speech data.

Furthermore, when considering the framework of convergent validity, which in various personality assessment instruments ranges from correlations of 0.29 to 0.57[52], our model's performance is deemed satisfactory. This comparison with convergent validity benchmarks, coupled with the ICC findings, reinforces the potential of voice-based personality assessments in providing consistent and reliable personality trait predictions.

Disattenuated correlations

Voice-based assessments, same as self-reports of personality, are susceptible to variabilities extrinsic to the personality traits they aim to measure, such as tonal fluctuations or ambient interference, potentially leading

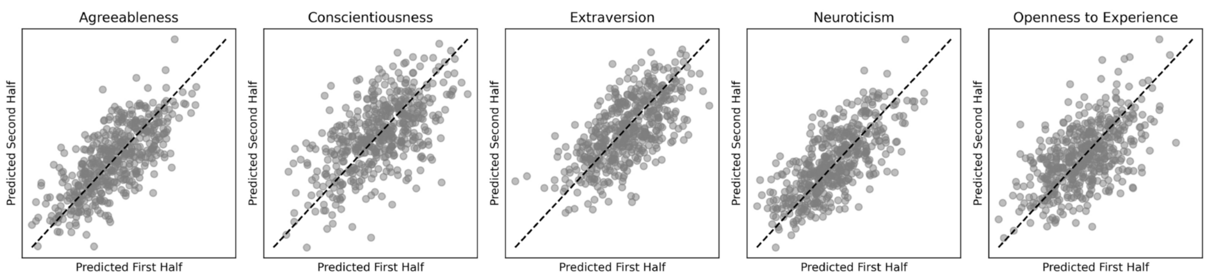

Figure 2. Scatterplots illustrating the intraclass correlation coefficients (ICC) for the Big Five personality traits, based on model predictions. Each plot compares the predictions from the first half of a speech recording against the second half, post-randomisation. The dashed line represents the line of perfect agreement. The density of data points along this line indicates moderate to high consistency of the model's predictions across different segments of speech.

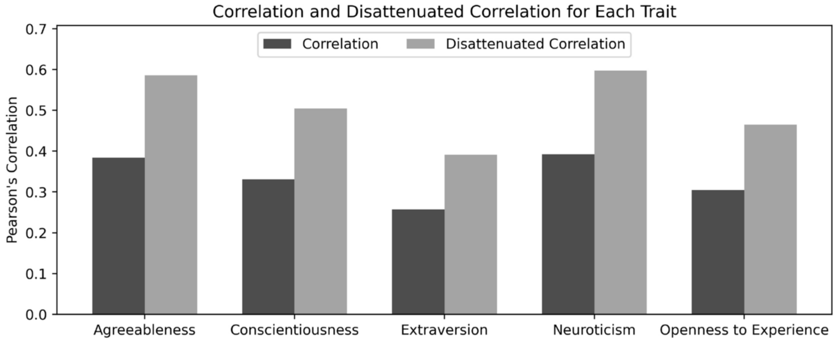

Figure 3. Bar graph presenting the comparison between the raw correlation coefficients and the disattenuated correlation coefficients for self-reported and model-predicted personality trait scores.

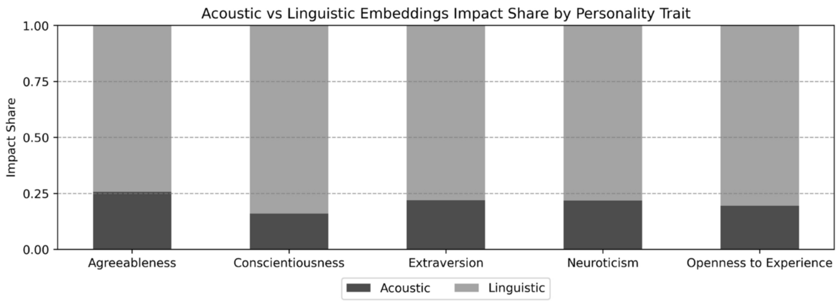

Figure 4. Stacked bar chart showing the normalized SHAP value impact shares of acoustic and linguistic embeddings on the prediction of Big Five personality traits. The graph demonstrates the varying degrees to which acoustic (tone, pitch) and linguistic (content, complexity) features contribute to the model's predictive accuracy for personality traits.

to attenuated correlation values. These manifest as measurement errors that lead to attenuated correlation coefficients between these instruments. To gauge the true correlation between voice- and questionnaire-based assessments, we use a correction with the aim of adjusting for measurement error, thereby offering a more accurate estimation of the true correlation between assessed traits and their vocal indicators. Using the dual correction formula[53,54], the disattenuated correlations between the predictions and self-reported scores range from 0.39 for extraversion to 0.60 for neuroticism. The range of disattenuated correlations observed, notably from moderate to high across different traits, not only validates the utility of voice-based assessments in capturing key personality dimensions but also paves the way for further methodological innovations in the field.

Differential contributions of acoustic and linguistic features

We sought to understand the relative impact of acoustic versus linguistic embeddings in predicting personality traits. As seen in Fig. 4, the impact share of acoustic and linguistic embeddings varies across the Big Five personality traits. Acoustic embeddings hold a larger share of impact on the prediction of agreeableness (26%), extraversion (22%) and neuroticism (22%), while acoustic embeddings have a comparatively lower influence on the prediction of openness to experience (19%), and conscientiousness (16%).

As shown in Fig. 4, extraversion, characterised by sociability and assertiveness, is notably influenced by acoustic features, possibly due to the vocal dynamics typically exhibited by extraverted individuals. Agreeableness, reflecting an individual's cooperative and friendly nature, also appears to be better captured through the tone and pitch of voice, hence the higher impact share of acoustic features. This aligns with the notion that agreeableness may manifest in more melodic and harmonious speech patterns, indicative of a more accommodating and consensus-seeking communication style. Similarly, neurotic individuals likely use a tone of voice that expresses anxiety or uncertainty.

On the other hand, linguistic embeddings play an important role in predicting conscientiousness, and openness. The content and complexity of language used by individuals can reveal the degree of their discipline (conscientiousness) and intellectual curiosity (openness). For instance, conscientious individuals may employ more organised and precise language, and those high in openness might demonstrate a richer vocabulary and the use of more abstract and imaginative language.

These findings suggest that while the acoustic qualities of speech are essential indicators of certain personality traits such as extraversion and agreeableness, the linguistic content provides critical insights into personality traits. This differential impact highlights the multifaceted nature of speech as a medium for personality

	A	C	E	N	O
Agreeableness	1.000	0.153	0.365	0.015	0.192
Conscientiousness	**0.063**	1.000	0.089	-0.321	0.081
	[-0.026; 0.152]				
Extraversion	0.408	**0.410**	1.000	-0.249	0.229
	[0.327; 0.490]	[0.329; 0.491]			
Neuroticism	**0.348**	**-0.641**	-0.275	1.000	-0.092
	[0.265; 0.432]	[-0.709; -0.572]	[-0.360; -0.189]		
Openness	0.196	**0.184**	**0.379**	-0.135	1.000
	[0.109; 0.283]	[0.096; 0.271]	[0.297; 0.461]	[-0.223; -0.046]	

Table 1. Observed and predicted trait intercorrelations with confidence intervals. The upper triangle of the matrix displays the intercorrelations among the Big Five personality traits based on self-reported scores, while the lower triangle presents the corresponding correlations derived from model predictions. The bolded figures in the lower triangle indicate statistically significant deviations from the self-reported correlations, with confidence intervals provided in brackets. Notably, the model's predicted correlations between Agreeableness and Conscientiousness, as well as between Openness and both Conscientiousness and Extraversion, align closely with the self-reported measures. However, significant discrepancies are evident in the negative correlation between Neuroticism and Conscientiousness, and to a lesser extent between Neuroticism and Agreeableness, underscoring areas for potential model refinement.

	True A	True C	True E	True N	True O
Predicted agreeableness	**0.38**	0.04	0.08	0.09	0.06
Predicted conscientiousness	-0.00	**0.36**	0.18	-0.26	0.05
Predicted extraversion	0.15	0.08	**0.26**	-0.13	0.14
Predicted neuroticism	0.16	-0.13	-0.17	**0.39**	-0.07
Predicted openness	0.12	0.02	0.12	-0.11	**0.29**
Multiple R	**0.19**	**0.29**	**0.21**	**0.36**	**0.17**

Table 2. Correlation matrix between predicted traits and self-reported traits, with multiple correlation coefficients (R) for non-targeted traits added as a new row. The diagonal values (in bold) represent the correlations between each predicted trait and its corresponding self-reported trait (convergent validity). Off-diagonal values represent correlations with non-targeted traits (discriminant validity). The bottom row shows the multiple R values for each predicted trait, indicating the combined correlation with all non-targeted self-reported traits; lower values suggest better discriminant validity.

assessment, where both what is said and how it is said contribute to a comprehensive understanding of the individual's personality profile.

Trait intercorrelations and discriminant validity

In the final analysis, trait intercorrelations derived from the model predictions in the test set were compared with the observed correlations from self-reported scores. This comparison revealed that our method broadly replicates the established correlation structure inherent in self-reported measures of personality. While all coefficients followed the same directional pattern, some variances were observed in specific instances. Notably, the correlations between agreeableness and conscientiousness, as well as between openness and conscientiousness, and openness and extraversion, approached the margins of the estimated confidence intervals. However, the model shows a deviation in capturing the traditionally low correlation between neuroticism and agreeableness, with an observed value of 0.015 against an estimated 0.348. Similarly, the correlation between neuroticism and conscientiousness was significantly exaggerated in the model's estimation, showing -0.641 as compared to the observed -0.321. All correlations are shown in Table 1, where correlations from the predicted traits are delineated in the lower triangle, with statistically significant deviations highlighted in boldface, and those from self-reported traits occupy the upper triangle.

To further assess the discriminant validity of our predictive models, we examined the correlations between each predicted personality trait and all self-reported traits, focusing on non-targeted traits. High discriminant validity is indicated when predicted traits correlate more strongly with their corresponding self-reported traits (demonstrating convergent validity) than with non-targeted traits.

Table 2 presents the Pearson correlation coefficients between each predicted trait and all self-reported traits. The diagonal values (in bold) represent the correlations between each predicted trait and its corresponding self-reported trait, demonstrating good convergent validity. The off-diagonal values show correlations with non-targeted traits, which are relevant for assessing discriminant validity.

We also calculated the multiple correlation coefficients (R) for each predicted trait with all non-targeted self-reported traits combined. These values quantify the extent to which non-targeted traits explain variance in the predicted traits (see Table 2). Lower multiple R values indicate better discriminant validity, as less variance in the predicted trait is explained by non-targeted traits.

The results show that each predicted trait correlates most strongly with its corresponding self-reported trait, supporting convergent validity. The correlations with non-targeted traits are generally lower, but some are notable. Predicted conscientiousness correlates negatively with self-reported neuroticism ($r = -0.26$) and positively with self-reported extraversion ($r = 0.18$), suggesting some overlap. Furthermore, predicted neuroticism shows moderate correlations with self-reported agreeableness ($r = 0.16$) and extraversion ($r = -0.17$), and has the highest multiple R value ($R = 0.36$), indicating that it may capture variance from other personality dimensions.

These findings suggest that while our models effectively capture trait-specific information (as evidenced by higher correlations on the diagonal), there is some shared variance with other traits, potentially due to overlapping speech features influencing multiple personality dimensions.

The higher intercorrelations among predicted traits compared to self-reported traits (as shown in Table 1) indicate that the models may be capturing shared variance across traits, which could affect discriminant validity. Potential improvements could be achieved by incorporating more trait-specific features or emphasising features that are unique to each trait or using feature selection methods to reduce overlap and eliminate features that contribute to shared variance among traits.

Having demonstrated the empirical findings of our investigation, encompassing the predictive accuracy, reliability and validity of our model for voice-based personality assessment, we aim to engage further with the results. In the forthcoming discussion, we will contextualize our findings within the landscape of psychological research and computational personality assessment, highlight the broader implications, examine prospective directions and challenges that lie ahead in the domain of computational personality assessment.

Discussion

This study sought to bridge the gap between speech characteristics and personality traits by employing advanced machine learning techniques to predict the Big Five personality dimensions from free-form speech samples. The findings contribute to the growing body of literature on computational personality assessment, demonstrating that both acoustic and linguistic features of speech can serve as significant indicators of personality traits.

The correlation coefficients between predicted and self-reported personality scores ranged from 0.26 to 0.39, with the highest correlations observed for neuroticism and agreeableness. These results are consistent with previous research suggesting that certain personality traits are more readily detectable through speech patterns than others[22,55]. The higher predictive accuracy observed for neuroticism and agreeableness suggests that these traits are more distinctly manifested in speech patterns. Neuroticism, characterized by emotional instability and heightened sensitivity to stress, may influence vocal expressions such as increased pitch variability, speech disfluencies, and changes in tone that reflect anxiety or tension[56]. These vocal cues can make neuroticism more detectable through acoustic analysis. Similarly, agreeableness, associated with empathy, cooperation, and social harmony, may be conveyed through warm vocal tones, friendly prosody, and positive language use[57,58]. These findings align with prior research indicating that emotional and interpersonal aspects of personality are often effectively communicated through vocal channels[57].

Despite the strong predictive accuracy for neuroticism, our model exhibited low discriminant validity for this trait, indicating that it was not entirely distinct from other personality dimensions in our predictions. This overlap suggests that the vocal cues associated with neuroticism, such as emotional expressiveness and variability, may also be present in the expression of other traits, leading to correlations across multiple personality dimensions. Similar challenges in differentiating neuroticism have been noted in previous studies, where the emotional aspects of neuroticism often intersect with features of other traits, complicating its distinct assessment through vocal analysis[59].

Interestingly, extraversion—the trait commonly believed to be most observable in social interactions—had the lowest correlation (0.26) between predicted and self-reported scores. This finding contrasts with several studies that have identified extraversion as one of the most detectable traits through vocal cues[22,36]. One possible explanation is that extraversion may be more context-dependent, with its vocal manifestations more pronounced in interactive or socially engaging settings rather than in monologues or self-introductions[60]. Moreover, the free-form nature of the speech samples may not have elicited the dynamic vocal expressions typically associated with extraverted behaviour. Future research could explore this by incorporating more interactive speech tasks or by examining the role of situational context in the expression of extraversion.

The study also revealed that linguistic features had a more substantial impact on the prediction of conscientiousness and openness, while acoustic features were more influential for agreeableness, extraversion, and neuroticism. This differential contribution aligns with theoretical expectations; conscientiousness and openness are often associated with cognitive and verbal expressions, such as the use of complex vocabulary and abstract concepts[21,61]. In contrast, traits like agreeableness and neuroticism are more closely linked to emotional states that can be conveyed through tone, pitch, and other acoustic properties[56].

The integration of both acoustic and linguistic features in the predictive models underscores the multifaceted nature of speech as a medium for personality expression. Previous studies have often focused on either acoustic or linguistic features in isolation[22,55], but our findings suggest that a combined approach yields a more comprehensive understanding. This aligns with the dual-channel hypothesis, which posits that both what is said (linguistic content) and how it is said (acoustic delivery) are crucial for conveying psychological information[62].

In terms of practical implications, while the results are promising, it is important to exercise caution before applying voice-based personality assessments in real-world settings. Ethical considerations are paramount, particularly regarding privacy and consent when recording and analysing individuals' speech[63]. Additionally, the legal implications of using such assessments, for example in employment or sales contexts, require careful scrutiny to ensure compliance with fast-evolving data protection regulations, such as GDPR, and to avoid potential biases or discrimination.

The study's limitations provide avenues for future research. First, the sample consisted predominantly of individuals from the United Kingdom, which may limit the generalisability of the findings across different cultures and languages. Cross-cultural studies could examine how cultural norms influence speech patterns and the expression of personality traits. Second, the speech samples were collected in a controlled, non-interactive setting, which may not capture the full range of vocal behaviours exhibited in naturalistic social interactions. Incorporating more dynamic and interactive speech tasks could enhance the ecological validity of the findings. Third, although the 50-item IPIP Big Five inventory is suitable for research purposes, incorporating longer and more comprehensive personality assessments in future studies could enhance the depth and reliability of the models, further validating their applicability in real-world settings.

Furthermore, while the models demonstrated moderate correlations with self-reported personality measures, the issue of criterion validity remains. It is not yet clear how well voice-based assessments predict real-world outcomes associated with personality traits, such as job performance, interpersonal relationships, or mental health indicators. Further studies examining the predictive validity of these assessments would be valuable.

In conclusion, this study contributes to the understanding of how speech reflects personality traits and demonstrates the potential of machine learning models in this domain. By highlighting both the possibilities and limitations, we encourage a balanced perspective that recognises the value of technological advancements while acknowledging the need for ethical considerations and further empirical validation. Voice-based personality assessment offers a promising complement to traditional methods, but its development and application should proceed with careful attention to scientific rigour and ethical responsibility.

Methods

Sample and procedure

The data were collected using Prolific in five separate rounds (November 2022, and February, April, June, July 2023). A total of 2525 respondents answered the survey call and started the survey by agreeing to informed consent prior to their participation in the study. Respondents could withdraw from the study at any moment and had the possibility to contact the research team via the platform to demand deletion of their data. Entire data collection was performed in accordance with relevant guidelines and regulations. Respondents were explicitly asked not to mention any personally identifying information in their speech recordings. Furthermore, no personal identifiable information was used for model training or evaluation. Respondents were paid higher than a minimum wage for their responses (£12 per hour) and the median duration of a submission was 10 min and 39 s. The data collection protocol was reviewed and approved by the advisory board.

In the survey, the respondent was asked to answer 50 questions from the International Personality Item Pool (known as IPIP), 10 questions per personality trait[64]. The questions are in mixed order, to avoid clear trait blocks, and are framed in positive and negative directions, to prevent acquiescence bias. A survey-based attention check was used by asking to what extent they agree with the following statement about themselves: "Swim the Atlantic to get to work". We filtered out respondents unless they answered "Somewhat disagree" or "Disagree". The theoretical model showed a good fit to the data, and all items showed statistically significant loadings on their theoretically expected scales. Descriptive statistics as well as the full list of questions are provided in the Supplementary Material.

The personality questionnaire was followed by a page with an audio recorder, where the respondents could record their speech sample. Respondents were instructed not to intentionally prepare a speech, and were shown very general prompts on what to speak about—e.g. "Introduce yourself!", "Tell us about your career to date.", and "Tell us about a challenge or conflict you've faced at work and how you managed it.", but it was made clear that following these was not mandatory and the respondents could speak freely (the full text can be found in the Supplementary Material). We introduced the prompts to alleviate the anxiety that tends to occur when respondents are asked to record their voice, with no intention of steering their responses into specific topics. Comprehensive checks were made on the collected audio data to ensure that no empty recordings or significantly distorted audio recordings made it into the training or test set. No filtering was done on the content of the audio recording to avoid restraining the diversity in the data.

The median duration of the speech audio recording was 190.9 s, ranging from 181.9 s and 223.3 s for the 25th and 75th percentiles respectively. Some outliers in terms of duration occurred, with the maximum duration of 998.8 s. We kept these observations in the sample as our results are robust even when removed from the sample. The median of words per recording was 406, with the length ranging between 331 and 499 for 25th and 75th percentiles respectively. The topics that were mostly discussed comprised careers, personal, problem-solving, aspirations, and job satisfaction. Duration of recordings, number of words, and extracted topics are visualised in Fig. 5.

The study's final sample consisted of $N = 2045$ participants, mirroring the UK's demographic profile in terms of gender, ethnicity, and primary language as per the 2021 census data (Fig. 6). Gender distribution was nearly balanced (51.3% female, 48.3% male, 0.3% undisclosed). In terms of ethnicity, the sample comprised 80.9% White, 8.5% Asian, 4.8% Black, 3.1% Mixed, and 1.4% identifying as Other. The majority (89.5%) reported English as their first language, with the remaining 10.5% English as their second or third language. The age range was skewed, excluding individuals below 18 years ($\bar{x}_{\text{age}} = 41.9$, $\sigma_{\text{age}} = 28.1$), to focus on adult populations.

Models

Our computational framework is founded upon a two-tiered model structure, designed to integrate acoustic and linguistic data for the prediction of the Big Five personality traits. The objective is to harness the distinct

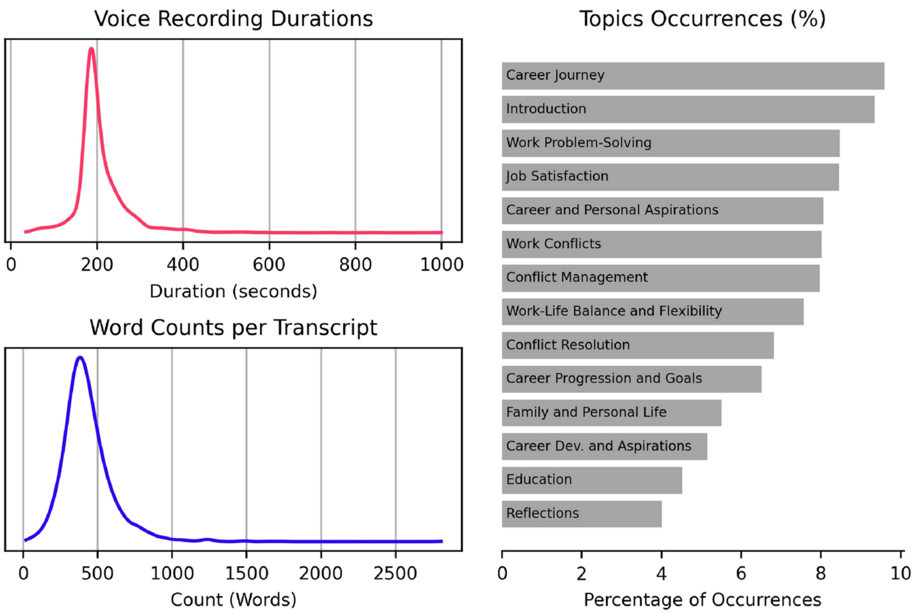

Figure 5. Analysis of speech samples from respondents. The top left graph represents the distribution of voice recording durations, showcasing a median duration of 190.9 s. The bottom left graph displays the word count distribution per transcript, with the median of 406 words. The bar chart to the right details the percentage of occurrence for various discussion topics, indicating a higher prevalence of career-related subjects and personal reflections among participants.

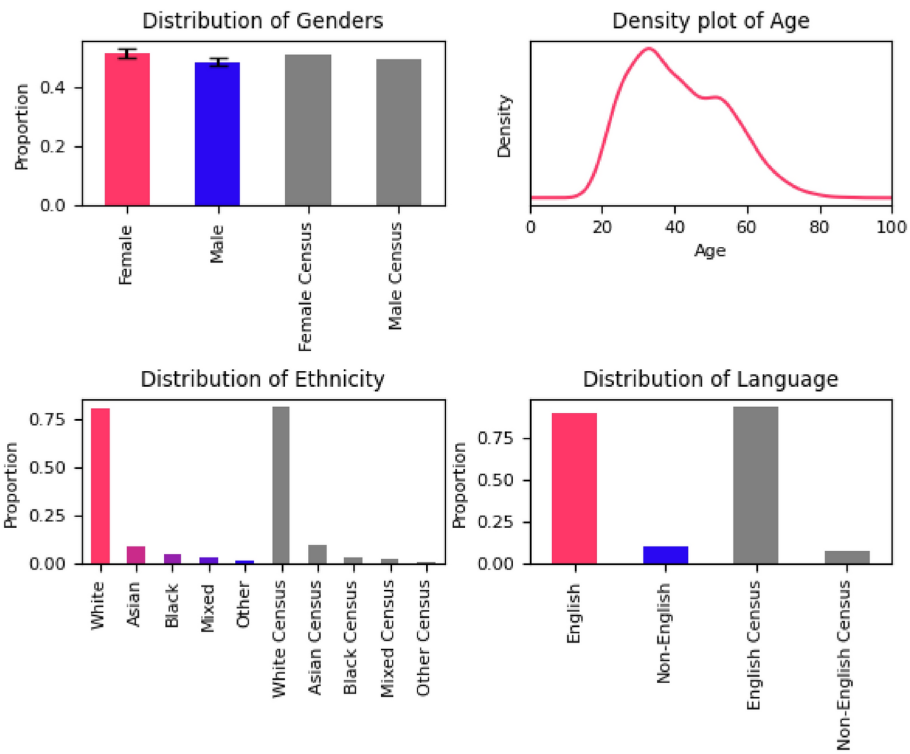

Figure 6. Demographic data of the sample compared to the 2021 UK Census ($N = 2045$).

yet complementary nature of speech signals—the acoustic and linguistic signals—to train a model capable of predicting an individual's personality profile.

Acoustic feature extraction

To extract the acoustic features of speech, we employed the YAMNet model[65]—a deep neural network designed for audio event detection. YAMNet is built upon the efficient MobileNet architecture, incorporating depth-wise separable convolutions to reduce complexity while maintaining high accuracy. It consists of 14 convolutional layers that are adept at capturing a wide range of acoustic characteristics across multiple sound events, making it highly suitable for processing complex speech signals. YAMNet was pre-trained on the AudioSet dataset, which comprises an expansive collection of over 2 million labeled 10-second sound clips drawn from YouTube videos. These clips span 521 audio event classes, providing a rich variety for the model to learn a wide range of acoustic features. The training process enables YAMNet to develop an understanding of various sound patterns, which is critical for generating robust and meaningful embeddings from audio inputs.

For our dataset, audio recordings were fed into YAMNet without modification to preserve the integrity of the natural speech patterns. The model's input requires a single-channel (monaural), 16-bit PCM-encoded audio with a 16 kHz sample rate. The audio waveform was normalised between −1 and 1 to ensure consistent amplitude levels. First, each recording, regardless of its original length, was transformed into a spectrogram using Short-Time Fourier Transform converting the time-domain signal into a frequency domain representation. Second, the model extracts log mel spectrogram features, which serve as the input into YAMNet. The network processes spectrograms through its convolutional layers, each layer building upon the previous one to extract increasingly abstract features. As the audio signal propagates through the network, the model captures various aspects of the sound.

The output from YAMNet's final layer is then subjected to mean pooling, which averages the embeddings across time, resulting in a consistent 1024-dimensional vector for each recording. This embedding encapsulates the salient acoustic information of the recording, which is crucial for the subsequent personality trait prediction tasks. The mean-pooled embedding vector serves as a compact representation of the audio's acoustic features, ready to be combined with linguistic embeddings for the comprehensive personality prediction model.

Linguistic feature extraction

For the linguistic feature extraction, our model capitalises on the capabilities of OpenAI's 'text-embedding-ada-002' model. This model has been trained to understand and convert text into numerical representations, capturing the essence of language in a form that machines can process for a variety of applications such as search, clustering, recommendations, anomaly detection, and classification.

The 'text-embedding-ada-002' model operates on input text of varying length—in our case the transcripts of recordings, generating a 1536-dimensional vector that encapsulates the semantic and syntactic properties of the language. This compact representation is achieved through a transformer architecture that can handle extended context lengths up to 8192 tokens, which is particularly beneficial for processing long documents without losing context, and pre-trained on trillions of examples of text. The embedding process begins by feeding transcriptions generated by Whisper, a state-of-the-art speech-to-text system, into the Ada model. The model then processes the textual input, translating it into a high-dimensional space where each dimension represents a learned feature of the language.

Data preparation and model training

In preparation for model training, we concatenated the 1024-dimensional acoustic embedding vectors and the 1536-dimensional linguistic embedding vectors, to form a single 2560-dimensional feature vector for each voice recording. The resulting composite vector encapsulates the acoustic and linguistic features of human speech, providing a foundation for subsequent predictive modelling. Our approach was guided by the hypothesis that the interactions between these vectors-representing both the content and manner of speech-would synergistically enhance model performance beyond what could be achieved by each set of features in isolation.

We used the XGBoost algorithm, an efficient implementation of gradient boosted trees, to train regression models for each of the Big Five personality traits (one for each trait). The selection of gradient boosted trees aimed to capitalise on the modelling of the interactions between acoustic and linguistic embedding vectors. To ensure the robustness and generalisability of the models, we employed a five-fold cross-validation strategy on the training dataset, which comprised 80% of the total sample. The hyperparameters were tuned through a randomised search within a predefined grid, optimising for model performance while preventing overfitting. The training objective was to minimize the Mean Squared Error (MSE), providing a direct measure of prediction accuracy by quantifying the variance between the predicted and actual trait scores. We also inspected the Mean Absolute Error (MAE), a metric preferred for its interpretability and resistance to outliers, which provides an average magnitude of errors in predictions. The evaluation of our models was based on the correlation between the predicted and ground truth scores in the test set, as the absolute Big Five trait prediction error is not as practically useful as their high correlation with the truth.

During the training of each fold, we used early stopping on a validation set (a subset of the training data) to prevent overfitting. Specifically, we monitored the validation loss and stopped training when the loss did not improve for a set number of rounds (patience parameter). This approach helps to halt training before the model begins to overfit the training data, thereby improving its ability to generalize to unseen data. Upon establishing the optimal hyperparameter settings, the models underwent a final estimation phase on the training set. Finally, models were evaluated on a hold-out test set, which accounted for 20% of the data.

Calculation of consistency and reliability

To ascertain the reliability and consistency of our model's predictions, we used the Intraclass Correlation Coefficient (ICC) as a statistical measure. ICC is commonly used to evaluate the degree of agreement or conformity among different measurements or raters. In the context of our study, this involved a unique application of ICC to assess how closely the model's predictions for personality traits aligned across different segments of the same recording.

Specifically, we divided the recordings in the test set into two halves and then randomised their order. This approach allowed us to simulate the variability inherent in independent rating scenarios, while maintaining a controlled environment to accurately assess the consistency of the model's predictions. We then applied the model to each half of the recordings to generate predictions for the personality traits. The ICC was calculated by correlating the predicted values from the two halves of the recordings. By comparing the model's predictions on randomly ordered segments of the same recording, we could effectively gauge the internal consistency of the model. This method mirrors the traditional use of ICC in reliability studies, where the agreement between different raters or measurements is assessed.

Feature contributions using SHAP

To assess the relative impact of acoustic and linguistic features on the prediction of personality traits, we applied SHAP (SHapley Additive exPlanations)[66] analysis to the XGBoost models trained for each trait. SHAP values, derived from cooperative game theory, offer insight into the contribution of each feature to the prediction made by the model.

For each personality trait, we began by extracting SHAP values from the trained XGBoost model. We focused on the magnitude of impact by considering the absolute SHAP values. This approach allowed us to quantify the influence of each feature set—acoustic and linguistic—separately. To understand the average influence of each feature set, we computed the mean of these absolute SHAP values. Further, we normalized the SHAP values by the number of features in each set. This normalisation allowed us to compare the impact of the two different-sized feature sets on an equal footing. We summed the normalized SHAP values for the acoustic and linguistic features and then computed their respective shares of the total impact. This gave us a proportional view of how much each feature set contributed to the model's predictions.

By calculating and comparing these SHAP values, we aimed to uncover how the *what* (linguistic content) and *how* (acoustic properties) of speech jointly influence personality trait predictions. This analysis provided an understanding of the relative importance of content and delivery in speech for personality assessment.

Data availibility

The data that support the findings of this study are not openly available due to their containing information that could compromise the privacy and consent of research participants. The original dataset potentially includes sensitive and personally identifiable information (PII) that was collected as part of a business operation. While all participants provided informed consent for the collection and analysis of their data for research purposes, explicit consent for public sharing of the data was not obtained. Therefore, to protect participant privacy and adhere to ethical guidelines, the data cannot be made publicly available. Details about the data and the procedures for data collection can be found in the Methods section of this paper. For further inquiries about the study or potential access to the data under strict privacy controls and for specific research purposes, interested researchers may contact the corresponding author.

Received: 27 March 2024; Accepted: 25 November 2024
Published online: 03 December 2024

References

1. Pisanski, K. & Bryant, G. A. The Evolution of Voice Perception. *Oxford University Press*[SPACE]https://doi.org/10.1093/oxfordhb/9780199982295.013.29 (2019).
2. Pear, T.H. *Voice and Personality*. (Chapman and Hall, 1931)
3. Johar, S. *Psychology of Voice*. 9–15. (Springer, 2016). https://doi.org/10.1007/978-3-319-28047-9_2 .
4. Marrero, Z.N.K., Gosling, S.D., Pennebaker, J.W., & Harari, G.M. Evaluating voice samples as a potential source of information about personality. *Acta Psychol.* **230**, 103740 (2022). https://doi.org/10.1016/j.actpsy.2022.103740.
5. Hickman, L. et al. Automated video interview personality assessments: Reliability, validity, and generalizability investigations. *J. Appl. Psychol.* **107**(8), 1323–1351. https://doi.org/10.1037/apl0000695 (2022).
6. Ramezani, M., Feizi-Derakhshi, M.-R., & Balafar, M.-A. Text-based automatic personality prediction using KGrAt-net: A knowledge graph attention network classifier. *Sci. Rep.* **12**(1) (2022). https://doi.org/10.1038/s41598-022-25955-z.
7. Kachur, A., Osin, E., Davydov, D., Shutilov, K., & Novokshonov, A. Assessing the big five personality traits using real-life static facial images. *Sci. Rep.* **10**(1) (2020). https://doi.org/10.1038/s41598-020-65358-6.
8. Woods, C., Luo, Z., Watling, D., & Durant, S. Twenty seconds of visual behaviour on social media gives insight into personality. *Sci. Rep.* **12**(1) (2022). https://doi.org/10.1038/s41598-022-05095-0.
9. Hoppe, S., Loetscher, T., Morey, S.A., & Bulling, A. Eye movements during everyday behavior predict personality traits. *Front. Hum. Neurosci.* **12** (2018). https://doi.org/10.3389/fnhum.2018.00105.
10. Singh, S. & Singh, W. AI-based personality prediction for human well-being from text data: A systematic review. *Multimed. Tools Appl.*[SPACE]https://doi.org/10.1007/s11042-023-17282-w (2023).
11. Kern, M. L., McCarthy, P. X., Chakrabarty, D. & Rizoiu, M.-A. Social media-predicted personality traits and values can help match people to their ideal jobs. *Proc. Natl. Acad. Sci.* **116**(52), 26459–26464. https://doi.org/10.1073/pnas.1917942116 (2019).
12. McCarthy, P.X., Gong, X., Braesemann, F., Stephany, F., Rizoiu, M.-A., & Kern, M.L. The impact of founder personalities on startup success. *Sci. Rep.* **13**(1) (2023). https://doi.org/10.1038/s41598-023-41980-y.
13. Stachl, C., Boyd, R.L., Horstmann, K.T., Khambatta, P., Matz, S.C., & Harari, G.M. Computational personality assessment. *Pers. Sci.* **2** (2021). https://doi.org/10.5964/ps.6115.

14. John, O. P. & Srivastava, S. The big five trait taxonomy: History, measurement, and theoretical perspectives. In *Handbook of Personality: Theory and Research* 2nd edn (eds John, O. P. et al.) 102–138 (Guilford Press, 1999).
15. Poropat, A. E. A meta-analysis of the five-factor model of personality and academic performance. *Psychol. Bull.* **135**(2), 322–338. https://doi.org/10.1037/a0014996 (2009).
16. Digman, J. M. Higher-order factors of the big five. *J. Pers. Soc. Psychol.* **73**(6), 1246–1256. https://doi.org/10.1037/0022-3514.73.6.1246 (1997).
17. McCrae, R. R. & Costa, P. T. Personality trait structure as a human universal. *Am. Psychol.* **52**(5), 509–516. https://doi.org/10.1037/0003-066x.52.5.509 (1997).
18. Costa, P. T. & McCrae, R. R. Personality in adulthood: A six-year longitudinal study of self-reports and spouse ratings on the NEO personality inventory. *J. Pers. Soc. Psychol.* **54**(5), 853–863. https://doi.org/10.1037/0022-3514.54.5.853 (1988).
19. Specht, J., Egloff, B. & Schmukle, S. C. Stability and change of personality across the life course: The impact of age and major life events on mean-level and rank-order stability of the big five. *J. Pers. Soc. Psychol.* **101**(4), 862–882. https://doi.org/10.1037/a0024950 (2011).
20. Roberts, B. W., Walton, K. E. & Viechtbauer, W. Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychol. Bull.* **132**(1), 1–25. https://doi.org/10.1037/0033-2909.132.1.1 (2006).
21. Yarkoni, T. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *J. Res. Pers.* **44**(3), 363–373. https://doi.org/10.1016/j.jrp.2010.04.001 (2010).
22. Mairesse, F., Walker, M.A., Mehl, M.R., & Moore, R.K. Using linguistic cues for the automatic recognition of personality in conversation and text. *J. Artif. Intell. Res.* **30**, 457–500 (2007). https://doi.org/10.1613/jair.2349.
23. Polzehl, T. *Speech-Based Personality Assessment*. 21–41. (Springer, 2015). https://doi.org/10.1007/978-3-319-09516-5_2 .
24. Demenko, G. Intonation processing for speech technology. *Speech Lang. Technol.* 14–15 (2012)
25. Koutsoumpis, A. & Vries, R. E. What does your voice reveal about you? Trait activation of voice characteristics and their relation with personality and communication styles. *Eur. J. Pers.* **35**(3), 392–413. https://doi.org/10.1002/per.2286 (2021).
26. Koutsoumpis, A. et al. Beyond traditional interviews: Psychometric analysis of asynchronous video interviews for personality and interview performance evaluation using machine learning. *Pers. Psychol.* **77**(1), 49–84. https://doi.org/10.1111/peps.12511 (2024).
27. Weisberg, Y. J., DeYoung, C. G. & Hirsh, J. B. Gender differences in personality across the ten aspects of the big five. *Front. Psychol.* **2**, 178 (2011).
28. Schmitt, D. P., Realo, A., Voracek, M. & Allik, J. Why can't a man be more like a woman? Sex differences in big five personality traits across 55 cultures. *J. Pers. Soc. Psychol.* **94**(1), 168–182 (2008).
29. Roberts, B. W. & Mroczek, D. Personality trait change in adulthood. *Curr. Dir. Psychol. Sci.* **17**(1), 31–35 (2008).
30. Munro, M. J. & Derwing, T. M. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Lang. Learn.* **45**(1), 73–97 (1995).
31. Paulhus, D. L. Socially desirable responding: The evolution of a construct. In *The Role of Constructs in Psychological and Educational Measurement* (eds Braun, H. I. et al.) 49–69 (Lawrence Erlbaum Associates, 2002).
32. Van Vaerenbergh, Y. & Thomas, T. D. Response styles in survey research: A literature review of antecedents, consequences, and remedies. *Int. J. Public Opin. Res.* **25**(2), 195–217 (2012).
33. Vazire, S. Who knows what about a person? the self-other knowledge asymmetry (SOKA) model. *J. Pers. Soc. Psychol.* **98**(2), 281–300 (2010).
34. Griffith, R. L. & Peterson, M. H. The nature of deception and applicant faking behavior. In *A Closer Examination of Applicant Faking Behavior* (eds Griffith, R. L. & Peterson, M. H.) 1–20 (Information Age Publishing, 2006).
35. Mehl, M. R., Gosling, S. D. & Pennebaker, J. W. Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *J. Pers. Soc. Psychol.* **90**(5), 862–877. https://doi.org/10.1037/0022-3514.90.5.862 (2006).
36. Scherer, K. R., London, H. & Wolf, J. J. The voice of confidence: Paralinguistic cues and audience evaluation. *J. Res. Pers.* **7**(1), 31–44. https://doi.org/10.1016/0092-6566(73)90030-5 (1973).
37. Stern, J., Schild, C., Jones, B.C., DeBruine, L.M., Hahn, A., Puts, D.A., Zettler, I., Kordsmeyer, T.L., Feinberg, D., Zamfir, D., Penke, L. & Arslan, R.C. Do voices carry valid information about a speaker's personality? *J. Res. Pers.* **92**, 104092 (2021). https://doi.org/10.1016/j.jrp.2021.104092.
38. Peer, E., Rothschild, D., Gordon, A., Evernden, Z. & Damer, E. Data quality of platforms and panels for online behavioral research. *Behav. Res. Methods* **54**(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3 (2021).
39. Weiss, K., Khoshgoftaar, T.M., & Wang, D. A survey of transfer learning. *J. Big Data* **3**(1) (2016). https://doi.org/10.1186/s40537-016-0043-6.
40. Hosna, A., Merry, E., Gyalmo, J., Alom, Z., Aung, Z., & Azim, M.A. Transfer learning: A friendly introduction. *J. Big Data* **9**(1) (2022). https://doi.org/10.1186/s40537-022-00652-w.
41. Kumar, S.A., Sasikala, S. et al. Parkinson's speech detection using yamnet. In *2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation* (*ICAECA*). 1–5 (2023). https://doi.org/10.1109/ICAECA56562.2023.10200704.
42. Lee, E., Kim, D., Min, C., & Kang, S. Automatic recognition of vocal reactions in music listening using smart earbuds: Poster abstract. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems. SenSys '20*. 665–666 (ACM, 2020). https://doi.org/10.1145/3384419.3430466 .
43. Harvey, M. *Acoustic Detection of Humpback Whales Using a Convolutional Neural Network* (2018). https://blog.research.google/2018/10/acoustic-detection-of-humpback-whales.html.
44. Mahmood, O. How ZSL uses ML to classify gunshots to protect wildlife (2020). https://cloud.google.com/blog/products/ai-machine-learning/how-zsl-uses-google-cloud-to-analyse-acoustic-data.
45. White, T. The fight against illegal deforestation with TensorFlow (2018). https://blog.google/technology/ai/fight-against-illegal-deforestation-tensorflow/
46. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., & Dean, J. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*(Burges, C.J., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K.Q. eds.). Vol. 26. (Curran Associates, Inc., 2013). https://proceedings.neurips.cc/paper_files/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf.
47. Tezgider, M., Yildiz, B., & Aydin, G. Text classification using improved bidirectional transformer. *Concurr. Comput. Pract. Exp.* **34**(9) (2021). https://doi.org/10.1002/cpe.6486.
48. Choi, S., Lee, H., Park, E., & Choi, S. Deep learning for patent landscaping using transformer and graph embedding. *Technol. Forecast. Soc. Change* **175**, 121413 (2022). https://doi.org/10.1016/j.techfore.2021.121413.
49. Phan, L.V., & Rauthmann, J.F. Personality computing: New frontiers in personality assessment. *Soc. Pers. Psychol. Compass* **15**(7) (2021). https://doi.org/10.1111/spc3.12624.
50. Connelly, B. S. & Ones, D. S. An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychol. Bull.* **136**(6), 1092–1122. https://doi.org/10.1037/a0021212 (2010).
51. Pace, V. L. & Brannick, M. T. How similar are personality scales of the "same" construct? A meta-analytic investigation. *Pers. Individ. Differ.* **49**(7), 669–676. https://doi.org/10.1016/j.paid.2010.06.014 (2010).
52. Laros, J.A., Souza Peres, A.J., Andrade, J.M., & Passos, M.F.D. Validity evidence of two short scales measuring the big five personality factors. *Psicol. Reflexão Crítica* **31**(1) (2018). https://doi.org/10.1186/s41155-018-0111-2.

53. Spearman, C. The proof and measurement of association between two things. *Am. J. Psychol.* **15**(1), 72. https://doi.org/10.2307/1412159 (1904).
54. Muchinsky, P. M. The correction for attenuation. *Educ. Psychol. Meas.* **56**(1), 63–75. https://doi.org/10.1177/0013164496056001004 (1996).
55. Mohammadi, G., & Vinciarelli, A. Automatic personality perception: Prediction of trait attribution based on prosodic features extended abstract. In *2015 International Conference on Affective Computing and Intelligent Interaction* (ACII). 484–490 (2015). https://doi.org/10.1109/ACII.2015.7344614.
56. Scherer, K.R. Vocal markers of emotion: Comparing induction and acting elicitation. *Comput. Speech Lang.* **27**(1), 40–58 (2013). https://doi.org/10.1016/j.csl.2011.11.003 (special issue on Paralinguistics in Naturalistic Speech and Language).
57. Scherer, K. R. Personality markers in speech. In *Social Markers in Speech* (eds Scherer, K. R. & Giles, H.) 147–209 (Cambridge University Press, 1979).
58. Berry, D. S. Accuracy in social perception: Contributions of facial and vocal information. *J. Pers. Soc. Psychol.* **61**(2), 298–307 (1991).
59. John, O. P. & Srivastava, S. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In *Handbook of Personality: Theory and Research* (eds Pervin, L. A. & John, O. P.) 102–138 (Guilford Press, 1999).
60. Vinciarelli, A. & Mohammadi, G. A survey of personality computing. *IEEE Trans. Affect. Comput.* **5**(3), 273–291. https://doi.org/10.1109/TAFFC.2014.2330816 (2014).
61. Hirsh, J. B. & Peterson, J. B. Personality and language use in self-narratives. *J. Res. Pers.* **43**(3), 524–527. https://doi.org/10.1016/j.jrp.2009.01.006 (2009).
62. Mehrabian, A. & Wiener, M. Decoding of inconsistent communications. *J. Pers. Soc. Psychol.* **6**(1), 109–114. https://doi.org/10.1037/h0024532 (1967).
63. Williams, M. L., Burnap, P. & Sloan, L. Towards an ethical framework for publishing twitter data in social research: Taking into account users' views, online context and algorithmic estimation. *Sociology* **51**(6), 1149–1168. https://doi.org/10.1177/0038038517708140 (2017) (**PMID: 29276313**).
64. Goldberg, L. R. et al. The international personality item pool and the future of public-domain personality measures. *J. Res. Pers.* **40**(1), 84–96. https://doi.org/10.1016/j.jrp.2005.08.007 (2006).
65. Team, T. YAMNet: A Pretrained Deep Learning Model for Sound Event Classification. https://tfhub.dev/google/yamnet/1 (2023)
66. Lundberg, S.M., & Lee, S.-I. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 4768–4777 (2017)

Acknowledgements
I would like to acknowledge the invaluable support of Thomas Sherwood, whose contributions were instrumental from the conception to the completion of this work. His insightful feedback, dedication, and collaboration have been essential throughout the entire process, making this publication possible. I want to thank Ken Benoit, Edgar Whitley, Sabine Benoit, Friedrich Geiecke, Stu Kennedy, Anna Bialas, and Olivia Lohmeyer for their invaluable feedback on my ideas and earlier versions of this article.

Declarations

Ethics statement
This study was reviewed and approved by the Koios Advisory Board. All participants provided informed consent before participating in the study. Participants were informed about the purpose of the data collection and were guaranteed confidentiality. Participants had the option to withdraw from the study at any time.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-81047-0.

Correspondence and requests for materials should be addressed to M.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2024, corrected publication 2025