Assessment of Personality Dimensions Across Situations Using Conversational Speech

Alice Zhang, *Student Member, IEEE,* Skanda Muralidhar, Daniel Gatica-Perez, *Member, IEEE,* and Mathew Magimai-Doss, *Member, IEEE*

***Abstract*—Prior research indicates that users prefer assistive technologies whose personalities align with their own. This has sparked interest in automatic personality perception (APP), which aims to predict an individual's perceived personality traits. Previous studies in APP have treated personalities as static traits, independent of context. However, perceived personalities can vary by context and situation as shown in psychological research. In this study, we investigate the relationship between conversational speech and perceived personality for participants engaged in two work situations (a neutral interview and a stressful client interaction). Our key findings are: 1) perceived personalities differ significantly across interactions, 2) loudness, sound level, and spectral flux features are indicative of perceived extraversion, agreeableness, conscientiousness, and openness in neutral interactions, while neuroticism correlates with these features in stressful contexts, 3) handcrafted acoustic features and non-verbal features outperform speaker embeddings in inference of perceived personality, and 4) stressful interactions are more predictive of neuroticism, aligning with existing psychological research.**

***Index Terms*—apparent personality perception, computational paralinguistics**

I. Introduction

MODELING human personality is fundamental to the development of affective computing systems capable of personalized interactions. Recent user studies have found that users are more engaged with and have greater trust in assistive technologies that reflect or adapt to their own personalities, thereby demonstrating the value of personality-aware systems [1], [2], [3]. Consequently, there is a growing field of research on automatic personality perception (APP), the task of inferring one's personality as perceived by external judges. Unlike automatic personality recognition (APR), which focuses on inferring self-reported personality traits, APP captures perceived personality and better reflects the cues that affective computing systems are designed to interpret.

Describing an individual's personality is a complex task. The Five-Factor Model of Personality (Big-5) is a widely accepted description of personality traits in five dimensions: Extraversion, Agreeableness , Conscientiousness, Neuroticism , and Openness to Experience [4]. A number of previous works have leveraged this description of personality to infer traits using modalities such as speech [5], [6], facial expressions [7], body language [8], and physiological signals [9]. The APP task is not new, as there exists a body of personality computing works [10]; however, existing works have mainly utilized datasets in which subjects are annotated in a single context. For instance, in the widely used Speaker Personality Corpus [5], [11], individuals are rated solely based on their behaviors in news broadcasts. Thus, analyses from these datasets assume that personalities are static and independent of external factors.

On the other hand, psychology and organizational behavior researchers have long recognized that personality expression can be situation dependent [12], [13], [14]. Experts in these fields have spent decades researching and debating the stability of traits across roles and environments as seen in the personality-situation debate. While researchers have since generally acknowledged an interactionist view in which both an individual's traits and the given situation influence expression of personality through behavior, affective computing models often lack mechanisms to adjust for context [13], [14].

Despite the importance of studying personality in relation to situational context, computational analyses of personality across situations remain limited. Addressing this gap, we utilize the UbImpressed dataset in which participants engaged in both job-related neutral and stressful conversations and received personality ratings for both interactions [15]. Given the conversational nature of this dataset, we aim to pose the question: *To what extent do conversational features (e.g., speech, non-verbal cues) explain perceived personality dimensions across varying contexts?* To answer this, we investigate the following research questions (RQ):

RQ1 Is there a significant difference in annotations of perceived personality of the same participants across two job-related conversation scenarios?

RQ2 How does the relationship between perceived personality and conversational features vary across contexts?

RQ3 How do conversational features from different interactions differ in their inference of perceived personality?

To our knowledge, while other researchers have studied constructs such as job performance on the two situations in this dataset, ours is the first to investigate perceived personality across two different situations in this dataset [16], [17].

The paper is structured as follows: section II discusses relevant work in affective computing and psychology, section III describes the dataset, section IV discusses the methodology for three analyses performed to study the research questions, section V presents results of each analysis, and section VI concludes the paper.

Alice Zhang is with Idiap Research Institute, Switzerland and The University of Texas at Austin, USA. (email: alice.zhang@austin.utexas.edu)

Skanda Muralidhar, Daniel Gatica-Perez, and Mathew Magimai-Doss are with Idiap Research Institute, Switzerland.

II. Related Work

A. Contextual Variability in Expression of Personality Traits

Researchers in psychology have long discussed the extent to which personality traits versus situational factors play a role in shaping behavior as demonstrated by the person-situation debate [12]. Proponents of the "person" perspective argue that individuals exhibit behaviors relatively similar to themselves and distinct from others, suggesting that underlying personality traits drive these behavioral patterns. In contrast, advocates from the "situation" perspective found that individuals' behaviors are relatively inconsistent across time and situation, arguing that personality traits do not exist and behaviors are influenced by the situation more than the individual's disposition [18].

While many resolutions have been proposed, a synthesis resolution acknowledging that both personality traits and situational factors influence expressed behavior is generally accepted [13], [14], [18]. With respect to an individual's behavior over long periods of time, traits predict behavior well and explain differences in behavior between people. With respect to momentary behavior, an individual's behavior is variable and traits may not strongly describe behaviors [19]. Thus, both perspectives are necessary for a full understanding of personality. However, existing affective computing works in inferring personality have mainly focused on personality as a static trait while psychology research has shown that situational factors also influence expressed personality. For more comprehensive computing solutions to infer personality, the situation in which the solution is employed also needs to be taken into consideration.

B. Explaining Contextual Variability in Personality Within Job-Related Settings

Two theories have been proposed to explain variations in personality-related responses across job-relevant contexts: the situation strength principle [20] and trait activation theory (TAT) [21]. The situation strength principle proposes that strong situations, defined as environments in which rules, structures or cues provide clear guidance on an individual's expected behavior, restrict expression of personality. Meanwhile, weak situations, in which there are fewer cues regarding expected behavior, allow for greater expression of personality [20], [22]. Complementing this theory, TAT states that individuals express specific traits when situations offer opportunities for a specific trait to be expressed. For instance, an individual with high extraversion may exhibit higher extroverted behavior when working in a sales role involving customer interaction.

Complementary to research around dimensions of personality, recent psychological research has given rise to frameworks, such as DIAMONDS [23] and CAPTIONS [24], that characterize situations along a set of dimensions (e.g., *complexity* - does the situation require deep thinking?). Furthermore, dimensions of situations have been found to correlate with dimensions of Big-5 personality. For instance, pleasant situations are positively correlated with extraversion, agreeability, and openness, thereby suggesting that individuals engage in more prosocial behaviors in positive situations [23], [24].

This body of psychological work illustrates the significance of studying personality within its situational context. Empirical findings show that contextualized personality measures tailored to specific environments yield higher predictive validity than generalized measures. For instance, Shaffer *et al.* [25] found that conscientiousness more strongly predicts performance in routine jobs, while its predictive power decreases in roles requiring higher levels of cognitive ability. Additional studies have shown similar findings that contextualized personality measures outperform non-contextualized personality measures in the inference of work-related creative problem solving [26], as well as job satisfaction and frustration [27].

Together, these findings motivate the need for computational approaches that account for situational variability when inferring personality, especially in job-related interactions.

C. Automatic Personality Perception using Speech

As speech signals encode rich information in addition to the spoken content itself, speech is a promising modality for the APP task. For instance, the 2012 INTERSPEECH Speaker Trait Challenge [11] resulted in a body of different approaches towards personality inference using speech ranging from low-level acoustic descriptors to spectrum analysis [28], [29]. More recent works showed that short utterance filler words [30] and dictionary learning of spectrograms [31] can be used to classify speaker's personality traits. However, prior works utilize datasets in which only one set of personality annotations was collected per participant, whether the data came from crowd-sourced monologue interview responses [6], [32], clips of video blogs from YouTube [33], [34], [35], or speech from news clips [5], [11], [28], [29] Therefore, these analyses assume that personalities are static and independent of spoken context.

In one of the few works that studies the relationship between variation in speech task and personality, Guidi *et al*. [36] showed that significant correlations between speech features and personality traits vary when reading neutral texts versus commenting on thematic apperception test images. Additionally, there are works that analyze personality and behavioral dimensions of speakers specifically in job interview settings [6], [32], [37]. However, these studies do not compare the speech or perceived personalities of the participants to their behaviors and personalities in other situations.

III. Dataset

We use the UbImpressed dataset, as it is the only dataset to our knowledge to contain personality annotations for the same participants across two different conversation scenarios [15]. Students at a hospitality school participated in a behavioral training program consisting of two lab sessions with performance feedback provided by professionals in human resources following the first lab session. Within each lab session, each student engaged in two dyadic role-play scenarios:

1) **Employment interview:** In this conversation, the student played the role of a mock applicant for a hospitality internship and a research assistant played the role of the interviewer. The interviewer asked the student questions regarding their motivation for a career in hospitality

TABLE I
INTRA-CLASS CORRELATION FOR PERCEIVED PERSONALITIES AND STRESS.

Session	Scenario	*ICC(2,k)*					
		Extra	Agree	Consc	Neuro	Open	Stress
1	Interview	0.65	0.54	0.59	0.51	0.48	0.74
	Desk	0.62	0.72	0.69	0.56	0.33	0.63
2	Interview	0.66	0.50	0.56	0.45	0.37	0.54
	Desk	0.77	0.58	0.59	0.37	0.58	0.45

and previous experiences. While job interviews can be stressful, this interview aimed to help students practice and did not carry the same weight as real interviews. We refer to this scenario as the *interview* scenario.

2) **Hotel reception desk interaction with an unsatisfied customer:** In this interaction, a research assistant played the role of a hotel client unsatisfied with charges on their bill and in a rush to resolve them. The student played the role of a receptionist working at the hotel front desk addressing the customer's complaints. The data collection protocol was designed such that this interaction was more hostile and thus more stressful for the student compared to the interview. The increased stress of this interaction is empirically verified in our analysis of annotations as will be shown in section V-A. We refer to this scenario as the *desk* scenario.

The dataset contains a total of 338 interactions evenly split between *interview* and *desk* interactions. The average duration of the *interview* interaction was 7.6 minutes and of the *desk* interaction was 4.6 minutes. 100 participants participated in the first lab session, and 69 participants returned to complete the second lab session. In this paper, *scenario* corresponds to the the *interview* or *desk* interaction and *session* corresponds to the first (1) or second (2) lab session of an interaction.

Perceived personality of participants was manually annotated for impressions of Big-5 personality traits and stress in both the *interview* and *desk* interaction. The *interview* and *desk* interactions were annotated by five and three independent annotators respectively. All annotators were Master's students in the same psychology program, thus providing a shared background for the annotation task. Each participant received a score on a scale of 1 to 7 for each dimension of personality and stress. Table I summarizes the agreement between raters, as assessed using the standard Intraclass Correlation Coefficient (ICC) measure of inter-rater reliability [38]. The agreement between all raters for most traits was greater than 0.5, indicating moderate reliability. For full details on data collection and annotation procedures, we refer to [15], [39].

IV. METHODS AND EXPERIMENTAL SETUP

A. Annotation Comparison across Conversations

To understand whether perceived personality differs across conversations (RQ1), we analyze the distribution of annotations across the *interview* and *desk* scenarios and the first and second sessions of each interaction. We employ the two-sample Kolmogorov-Smirnov (KS) test for goodness of fit where the null hypothesis states that the underlying continuous distributions $F(x)$ and $G(x)$ of two independent samples are identical for all x [40]. We first use the two-sample KS test to compare the distributions of participants' perceived stress across different scenarios and sessions to verify that the two interaction scenarios in the UbImpressed dataset are significantly different in order to validate the use of the UbImpressed dataset to study our research questions. Then, we use the two-sample KS test to compare the distribution of perceived personality across different scenarios and sessions to answer RQ1.

B. Feature Extraction and Selection

We diarize each conversation and retain only the audio segments corresponding to speech from the students who played the role of the job applicant and hotel receptionist. From each student's speech, we set to extract features that have been shown to vary with a speaker's emotional state and influenced by personality traits [41]. Consequently, we choose features previously validated for speech emotion recognition and investigate their potential to infer personality. Specifically, we extract three sets of features: (1) eGeMAPS features, (2) speaker embeddings, and (3) non-verbal features.

eGeMAPS [42] features are a set of acoustic features commonly used for speech-based affective computing using the openSMILE toolkit [43]. It contains 88 features that capture frequency, energy, amplitude, and spectral parameters initially hand-crafted for speech emotion recognition.

Speaker embeddings are fixed-dimensional representations of speech, such as x-vectors [44], that encode speaker identity. ECAPA-TDNN vectors [45] further improve upon the time-delay neural network architecture from which x-vectors are extracted, and Ulgen *et al.* [46] showed that intra-speaker ECAPA-TDNN embedding clusters reveal emotion states. Thus, we focus on 512-dimensional ECAPA-TDNN vectors extracted via the Pyannote toolkit [47].

Non-verbal features include audio and visual cues selected for their relevance in existing literature in psychology and social computing spanning five categories for a total of 75 features previously extracted in the UbImpressed dataset [15]. The categories include: speaking activity features (21 dimensions), prosody features (30 dimensions), head nods (8 dimensions), visual back-channeling (6 dimensions) and overall visual motion features (10 dimensions).

For the eGeMAPS and speaker embedding features, we extract one feature vector from each student's turn (i.e. utterance) within each scenario and session. We explore aggregation of the utterance-level features via a median and a mean operation such that each dimension of the feature vector is represented by the median or mean of the feature dimension across all of the participant's utterances in a specific conversation. Through this aggregation, for a given scenario and session, there is only one feature vector per participant. We found that personality inference using the median feature vector representation outperformed the alternatives (mean feature vector and one feature vector per utterance) and focus on methods and results obtained using the median feature vector onward.

On the other hand, the non-verbal feature set already includes aggregated statistics of non-verbal features over the participant's entire interaction within a scenario and session. Therefore, there is only one feature vector extracted for each participant per interaction, and we use the feature vector as is.

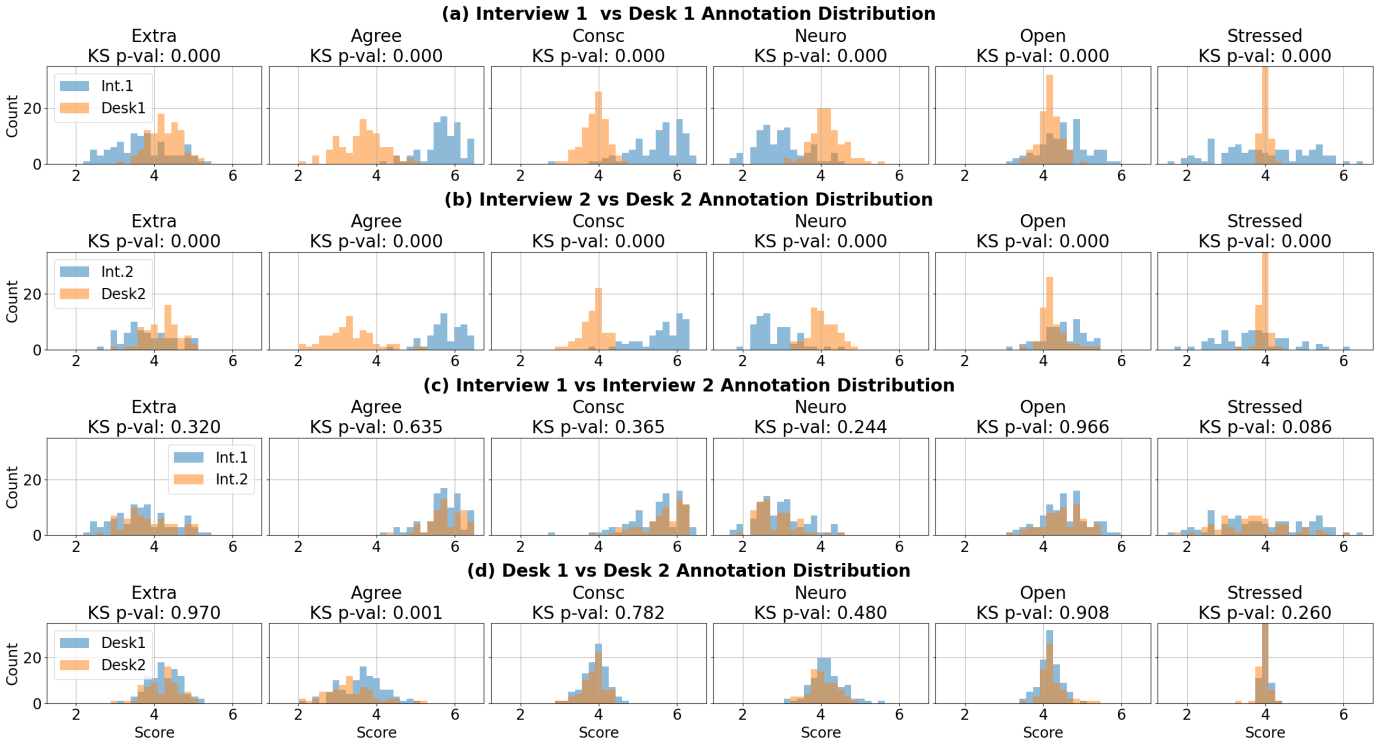

Fig. 1. (RQ1) Comparing distribution of perceived personality scores between: (a) interview and desk scenario of first lab session, (b) interview and desk scenario of second lab session, (c) interview scenario of first and second lab sessions, and (d) desk scenario of first and second lab sessions.

Lastly, for each feature set, we select a subset of features using the Pearson Correlation Coefficient (PCC). The PCC measures the strength of the linear relationship between two continuous variables. We assume that relevant information is contained in features significantly correlated with perceived personality scores and select features with $p < 0.05$ (RQ2).

C. Experimental Setup for Inference of Perceived Personality

We formulate the inference of perceived personality from speech signals as a regression task (RQ3). To compare personality inference across conversations, we do not discretize personality scores into "high" and "low" categories relative to mean or median annotations of the conversation.

1) Conversation-Specific Inference: First, we aim to infer perceived personality on a specific conversation scenario and session to understand how feature selection impacts performance. We perform feature selection via the method described in section IV-B using data from the complementary session of the same scenario. For instance, to build and evaluate a system to infer perceived personality on the first *interview* session, we perform feature selection using data from the second *interview* session so that no data from the test set influences the feature selection process. Then, we train a Random Forest (RF) Regressor with 100 trees to infer a score for each personality dimension using one of the three feature sets or a subset of the feature sets. We evaluate all regressors using the coefficient of determination R^2, which reports the amount of total variance explained by the model, and the PCC between the predicted and observed personality scores in a 10-fold cross validation scheme. As each feature vector corresponds to one participant, each evaluation fold consists of speakers unseen in the training process. Thus, the evaluation is speaker independent. We obtain confidence intervals by bootstrapping the evaluation set 100 times for each iteration of evaluation.

2) Inference across Scenarios and Sessions: To understand the generalization of personality regression across different scenarios within the same set of participants, we build a RF regressor trained on data from one scenario (e.g. *interview 1*) and evaluate the regressor using data from the complementary scenario of the same session (e.g. *desk 1*). We perform feature selection using the same method as in sections IV-B and IV-C1. Then, to understand how a trained RF regressor generalizes across conversation instances of the same scenario within a set of participants, we build a regressor trained on data from one session (e.g. *interview 1*) and evaluate the regressor using data from the complementary session of the same scenario (e.g. *interview 2*). In this setup, we perform feature selection using the training dataset.

V. Results and Analysis

A. Annotation Comparison Results and Analysis

As shown in Figure 1, with a p-value threshold of 0.05, we reject the null hypothesis of the two-sample KS test in favor of the alternative hypothesis when comparing all personality dimensions and stress scores across *interview* and *desk* scenarios. In contrast, we fail to reject the null hypothesis when comparing personality dimensions and stress across the first and second sessions of the *interview* or *desk* scenario, with the exception of agreement in the *desk* scenario. First, this analysis shows that participants are perceived to exhibit statistically significant higher stress in the client interaction (mean stress $= 4.06$) compared to the interview (mean stress $=$

3.82) and validates the suitability of the UbImpressed dataset to study our research questions.

Overall, this analysis shows that perceived personality differs significantly between the neutral interview and stressful client interaction for the same participants (RQ1). In the *interview* scenario, participants were perceived as more agreeable (mean = 5.81 in *interview* vs. 3.53 in *desk*) and open (mean = 4.52 in *interview* vs. 4.19 in *desk*), and less neurotic (mean = 2.92 in *interview* vs. 4.11 in *desk*), which agrees with psychology studies on dimensions of situations discussed in section II-B. These studies found that positive situations are positively correlated with agreement and openness [23], [24], which aligns with the *interview* scenario that was designed to allow students to practice for future interviews and was therefore more relaxed and decontracted compared to the *desk* interaction. This observation also aligns with TAT, which suggests that stressful situations can trigger the expression of neuroticism and make it more observable, leading to higher perceived neuroticism in the client interaction.

B. Correlation Analysis Results and Analysis

TABLE II
(RQ2) CORRELATION OF SELECT EGEMAPS AND NON-VERBAL FEATURES WITH PERSONALITY. † $p < 0.01$; * $p < 0.05$.

Feature	Scenario	Extra	Agree	Consc	Neuro	Open
equivalent Sound Level (dB) (eGeMAPS)	Int. 1	0.52†	0.43†	0.31†	-0.08	0.41†
	Int. 2	0.41†	0.34*	0.37†	-0.27*	0.27*
	Desk 1	0.09	0.41†	0.04	-0.36†	0.19
	Desk 2	0.21	0.49†	0.15	-0.27*	0.07
spectralFluxV-sma3nz-amean (eGeMAPS)	Int. 1	0.51†	0.40†	0.24*	-0.07	0.39†
	Int. 2	0.46†	0.31*	0.33*	-0.21	0.29*
	Desk 1	0.06	0.44†	0.08	-0.42†	0.23†
	Desk 2	0.18	0.46†	0.15	-0.37†	-0.07
# pauses >1 second (non-verbal)	Int. 1	0.18	0.23*	0.16	-0.07	-0.27*
	Int. 2	0.31*	0.22	0.04	-0.13	0.38†
	Desk 1	-0.12	0.23*	-0.03	-0.27*	0.24*
	Desk 2	0.07	0.15	0.07	-0.02	0.02
Count Head Nods (non-verbal)	Int. 1	0.42†	0.25*	0.24*	-0.06	0.31†
	Int. 2	0.44†	0.01	-0.07	-0.07	0.29*
	Desk 1	0.25*	0.07	0.10	-0.24*	0.08
	Desk 2	0.39†	0.22	0.04	-0.30*	0.09

Select results of the correlation between each participant's median eGeMAPS feature vector and non-verbal feature vector, and personality dimensions are presented in Table II (RQ2). While we also compute the correlation between each dimension of the speaker embedding and personality, speaker embeddings are not interpretable in the manner eGeMAPS and non-verbal features are; therefore, we omit the results from the table. The complete correlation matrix for all feature sets can be found here[1].

For both sessions, there are stronger correlations between eGeMAPS and nonverbal features and personality in the *interview* scenario for all personality dimensions except neuroticism, in which the correlation between neuroticism and features is stronger in the *desk* scenario. Additionally, we observe an inverse correlation between neuroticism and features compared to the correlation of the other four personality dimensions. We hypothesize this is because out of the five personality dimensions, neuroticism is the only dimension associated with negative emotions such as anxiety, and prior research has shown that anxiety introduces irregularities in the produced speech and nonverbal behaviors compared to positive or neutral speech [48], [49].

For eGeMAPS features, we observe a common subset of features that are significantly correlated with all personality dimensions except neuroticism in the *interview* scenario. Energy- (loudness and equivalent sound level) and spectral- (spectral flux) related features are all correlated with perceived extraversion, agreement, conscientiousness, and openness in the *interview*. In contrast, the correlation is significant between this subset of features and neuroticism only in the *desk* scenario. These results indicate that the correlation between speech features and neuroticism is more pronounced in stressful contexts. Interestingly, significant correlations between this subset of speech features and agreement exist in both the *interview* and *desk* interaction, suggesting that this trait may be more consistently expressed across contexts.

For non-verbal features, we generally observe that features related to speech energy, voicing rate, and head nods are positively correlated with all personality dimensions except neuroticism in the *interview* scenario. In the *desk* scenario, we observe positive correlations between features related to speaking activity, number of pauses, and speech energy, and agreement. On the other hand, this set of non-verbal features is negatively correlated with neuroticism in the *desk* interaction. Interestingly, the observed negative correlations between speech pauses and neuroticism differ from psychology research that found longer and more frequent speech pauses in the public speaking of anxious individuals [50], [51].

C. Inference of Perceived Personality Results

1) Conversation-Specific Perceived Personality: We observe that the RF regressor contains some predictive power for all personality dimensions using all three feature sets as seen in Table III. Generally, eGeMAPS and non-verbal features outperform speaker embedding features, with non-verbal features explaining a maximum 37% of the variance for the extraversion dimension. Our eGeMAPS results are comparable to the results obtained by Barchi *et al.* who report R^2 values between 0.12 and 0.22 for each personality dimension using eGeMAPS features and a dataset of YouTube videos [33]. To the best of our knowledge, there are no comparable works on personality regression using non-verbal and speaker embedding features. Closest similar works come from Luo *et al.* [52] who use x-vectors and a support vector regressor to infer neuroticism scores in healthy and depressed speech, achieving a R^2 of 0.22 and r of 0.49, and from Mawalim *et al.* [53] who use x-vectors and non-verbal cues to perform binary classification of personality dimensions.

Our regressor performs comparably with and without feature selection, showing that a subset of features can have as much predictive power as the entire feature set. Additionally, across both sessions and all feature sets, we observe that the *desk* dataset is better for inference of neuroticism and that the *interview* dataset is better for inference of the other four personality dimensions. These results align with the correlation results presented in section V-B and also psychological research on the inference of personality from

[1]https://osf.io/8vztd/?view_only=57e6b2ae1602457a89bf9472889a4d64

TABLE III

(RQ3) REGRESSION RESULTS FOR PREDICTING PERSONALITY DIMENSIONS WITHIN A GIVEN SCENARIO AND SESSION. WHERE THE DIMENSION OF FEATURES IS 0, THERE ARE NO STATISTICALLY SIGNIFICANT CORRELATIONS BETWEEN THE FEATURE SET AND THE PERSONALITY DIMENSION. ROWS IN BOLD HIGHLIGHT THE BEST PERFORMANCE FOR THE CORRESPONDING PERSONALITY DIMENSION ACROSS BOTH SCENARIOS. $^{\dagger}p < 0.01$; $^{*}p < 0.05$.

Feature	Session	Extraversion		Agreement		Conscientiousness		Neuroticism		Openness	
		Dims	R^2, r	Dims	R^2, r	Dims	R^2, r	Dims	R^2, r	Dims	R^2, r
Hand-crafted (eGeMAPS)	1	88	0.20 ± 0.20, 0.48†	88	0.16 ± 0.14*, 0.41†	88	-0.03 ± 0.13, 0.17	88	-0.11 ± 0.15, -0.02	88	0.06 ± 0.18, 0.29†
		27	0.21 ± 0.20, 0.49†	52	0.16 ± 0.16*, 0.42†	39	-0.13 ± 0.16, 0.08	15	-0.13 ± 0.16, 0.01	18	0.16 ± 0.18, 0.40†
	2	88	0.10 ± 0.18, 0.38†	**88**	**0.22 ± 0.17*, 0.52†**	**88**	**0.18 ± 0.19, 0.46†**	88	-0.14 ± 0.15, 0.07	88	-0.08 ± 0.16, 0.12
		31	0.04 ± 0.20, 0.33†	43	0.21 ± 0.20*, 0.51†	28	0.16 ± 0.19, 0.45†	6	-0.13 ± 0.15, 0.07	32	-0.08 ± 0.16, 0.12
Speaker Embeddings (ECAPA-TDNN)	1	512	-0.04 ± 0.08, 0.04	512	0.08 ± 0.08, 0.30†	512	-0.05 ± 0.12, 0.07	512	-0.21 ± 0.07, -0.35†	512	-0.06 ± 0.07, 0.01
		40	0.04 ± 0.13*, 0.24*	99	0.10 ± 0.10, 0.31†	64	0.02 ± 0.13, 0.23*	43	-0.17 ± 0.09, -0.17	20	0.01 ± 0.11, 0.20
	2	512	-0.11 ± 0.13, -0.03	512	0.15 ± 0.14*, 0.44†	512	0.02 ± 0.14, 0.24	512	0.10 ± 0.12, 0.36†	512	-0.14 ± 0.10, -0.16
		63	-0.09 ± 0.13, 0.09	61	0.17 ± 0.14, 0.45†	40	0.17 ± 0.15, 0.45†	13	-0.20 ± 0.17, -0.02	38	0.02 ± 0.12, 0.23
Non-verbal	1	75	0.35 ± 0.14†, 0.60†	75	0.15 ± 0.17, 0.38†	75	0.08 ± 0.18, 0.33†	75	-0.05 ± 0.12, 0.11	**75**	**0.25 ± 0.12*, 0.49†**
		35	**0.37 ± 0.16†, 0.61†**	23	0.21 ± 0.16*, 0.45†	15	0.01 ± 0.19, 0.25*	2	-0.24 ± 0.20, -0.11	18	0.11 ± 0.17, 0.36*
	2	75	0.28 ± 0.17*, 0.55†	75	0.07 ± 0.18, 0.35†	75	0.01 ± 0.23, 0.29*	75	-0.07 ± 0.16, 0.08	75	0.11 ± 0.15, 0.35†
		32	0.26 ± 0.19*, 0.53†	35	0.05 ± 0.18, 0.34†	28	0.04 ± 0.21, 0.32†	6	-0.14 ± 0.16, -0.05	36	0.11 ± 0.19, 0.36†

(a) Interview

Feature Type	Session	Extraversion		Agreement		Conscientiousness		Neuroticism		Openness	
		Dims	R^2, r	Dims	R^2, r	Dims	R^2, r	Dims	R^2, r	Dims	R^2, r
Hand-crafted (eGeMAPS)	1	88	-0.08 ± 0.12, -0.03	88	-0.02 ± 0.14, 0.20	88	-0.23 ± 0.13, -0.19	88	-0.03 ± 0.16, 0.21*	88	-0.08 ± 0.10, 0.04
		20	-0.03 ± 0.10, 0.11	26	-0.01 ± 0.17, 0.24*	0	-	13	0.07 ± 0.16, 0.33†	1	-0.22 ± 0.12, -0.13
	2	88	-0.07 ± 0.18, 0.15	88	0.07 ± 0.17, 0.34†	88	-0.28 ± 0.13, -0.43†	88	-0.03 ± 0.13, 0.12	88	-0.10 ± 0.15, 0.06
		5	0.03 ± 0.14, 0.25	24	0.14 ± 0.20, 0.42†	2	-0.31 ± 0.22, -0.13	21	-0.05 ± 0.15, 0.15	11	-0.15 ± 0.14, -0.03
Speaker Embeddings (ECAPA-TDNN)	1	512	-0.06 ± 0.10, 0.0	512	-0.01 ± 0.09, 0.15	512	-0.21 ± 0.10, -0.30†	512	-0.07 ± 0.09, 0.03	512	-0.03 ± 0.09, 0.08
		52	-0.09 ± 0.12, 0.03	50	0.01 ± 0.11, 0.20*	17	-0.16 ± 0.16, 0.0	44	-0.05 ± 0.13, 0.12	30	0.03 ± 0.10, 0.23*
	2	512	-0.07 ± 0.18, 0.15	512	0.07 ± 0.17, 0.34†	512	-0.28 ± 0.13, -0.43†	512	-0.03 ± 0.13, 0.12	512	-0.1 ± 0.15, 0.06
		38	-0.04 ± 0.16, 0.13	84	0.06 ± 0.13, 0.28*	10	-0.08 ± 0.22, 0.12	**48**	**0.17 ± 0.16, 0.43†**	29	-0.12 ± 0.12, -0.04
Non-verbal	1	75	0.01 ± 0.13, 0.19	75	0.04 ± 0.11, 0.26†	75	0.02 ± 0.13, 0.23*	75	0.03 ± 0.14, 0.25*	75	-0.19 ± 0.13, -0.14
		10	-0.08 ± 0.14, 0.12	16	-0.07 ± 0.17, 0.21*	1	-0.27 ± 0.18, -0.15	13	-0.19 ± 0.17, -0.01	2	-0.20 ± 0.17, -0.08
	2	75	-0.04 ± 0.19, 0.14	75	-0.15 ± 0.15, -0.03	75	-0.24 ± 0.12, -0.32*	75	-0.04 ± 0.20, 0.18	75	-0.21 ± 0.13, -0.17
		13	0.03 ± 0.21, 0.26*	30	-0.15 ± 0.18, 0.02	3	-0.23 ± 0.22, -0.08	28	0.04 ± 0.20, 0.28*	7	-0.29 ± 0.20, -0.29*

(b) Desk

TABLE IV

(RQ3) REGRESSION RESULTS FOR PREDICTING PERSONALITY DIMENSIONS WHEN TRAINING ON ONE SCENARIO AND EVALUATING ON THE OTHER SCENARIO WITHIN THE SAME SESSION. WHERE THE DIMENSION OF FEATURES IS 0, THERE ARE NO STATISTICALLY SIGNIFICANT CORRELATIONS BETWEEN THE FEATURE SET AND THE PERSONALITY DIMENSION. $^{\dagger}p < 0.01$; $^{*}p < 0.05$.

Session	Train Scenario	Eval. Scenario	Extraversion		Agreement		Conscientiousness		Neuroticism		Openness	
			Dims	R^2, r	Dims	R^2, r	Dims	R^2, r	Dims	R^2, r	Dims	R^2, r
1	Interview	Desk	88	-0.41 ± 0.29, 0.17	88	-19.1 ± 5.0, 0.11	88	-32.0 ± 8.1, 0.07	88	-8.6 ± 2.8, 0.12	88	-7.7 ± 2.0, 0.14
			27	-0.64 ± 0.34, 0.11	52	-19.9 ± 5.3, 0.13	39	-32.3 ± 7.9, -0.01	15	-8.2 ± 2.4, -0.11	18	-8.5 ± 2.3, 0.22*
	Desk	Interview	88	-0.47 ± 0.26, 0.27†	88	-20.5 ± 5.6, 0.31†	88	-6.0 ± 2.1, 0.07	88	-4.8 ± 1.7, 0.01	88	-0.4 ± 0.22, 0.25*
			20	-0.40 ± 0.28, 0.40†	26	-20.5 ± 5.6, 0.22*	0	-	13	-4.7 ± 1.6, 0.13	1	-0.33 ± 0.22, 0.17
2	Interview	Desk	88	-0.55 ± 0.43, 0.3*	88	-20.9 ± 8.8, 0.21	88	-41.6 ± 11.5, -0.01	88	-16.1 ± 4.7, -0.07	88	-2.6 ± 1.3, -0.11
			31	-0.75 ± 0.50, 0.29*	43	-20.9 ± 8.9, 0.22	28	-36.4 ± 10.0, 0.13	6	-17.0 ± 5.0, -0.04	32	-3.1 ± 1.4, 0.02
	Desk	Interview	88	-0.09 ± 0.15, 0.29*	88	-35.5 ± 10.9, 0.04	88	-11.8 ± 4.5, -0.21	88	-7.33 ± 3.4, -0.03	88	-0.40 ± 0.28, 0.01
			5	-0.29 ± 0.22, 0.10	24	-35.2 ± 10.8, 0.33†	2	-10.9 ± 4.1, -0.02	21	-7.3 ± 3.4, -0.08	11	-0.43 ± 0.29, 0.16

(a) eGeMAPS

Session	Train Scenario	Eval. Scenario	Extraversion		Agreement		Conscientiousness		Neuroticism		Openness	
			Dims	R^2, r	Dims	R^2, r	Dims	R^2, r	Dims	R^2, r	Dims	R^2, r
1	Interview	Desk	75	-0.10 ± 0.24†, 0.26*	75	-15.6 ± 3.8, 0.35†	75	-25.7 ± 6.6, 0.02	75	-6.7 ± 2.3, 0.12	75	-7.0 ± 1.8, 0.22*
			35	-0.30 ± 0.37†, 0.28*	23	-18.0 ± 4.0, 0.31†	15	-38.5 ± 9.6, 0.09	2	-9.9 ± 3.4, 0.19	18	-7.1 ± 1.8, 0.15
	Desk	Interview	75	-0.21 ± 0.17†, 0.38†	75	-16.8 ± 4.6, 0.34†	75	-6.5 ± 2.3, -0.16	75	-3.6 ± 1.3, 0.12	75	-0.47 ± 0.28, 0.30*
			10	-0.55 ± 0.31, 0.31†	16	-17.4 ± 4.9, 0.42†	1	-6.0 ± 2.2, 0.10	13	-4.31 ± 1.5, 0.28†	2	-0.22 ± 0.16, -0.17
2	Interview	Desk	75	-0.60 ± 0.50, 0.27*	75	-22.7 ± 9.1, 0.31*	75	-41.6 ± 11.0, -0.11	75	-17.3 ± 4.7, 0.11	75	-2.9 ± 1.4, -0.01
			32	-0.42 ± 0.47, 0.41†	35	-22.7 ± 9.2, 0.23	28	-39.4 ± 9.9, -0.01	6	-17.1 ± 4.6, -0.16	36	-2.2 ± 1.1, 0.12
	Desk	Interview	75	0.08 ± 0.15†, 0.48*	75	-31.6 ± 9.9, -0.02	75	-10.5 ± 4.1, 0.07	75	-7.17 ± 3.4, 0.13	75	-0.21 ± 0.18, -0.06
			13	-0.10 ± 0.25, 0.45†	30	-32.8 ± 10.2, 0.06	3	-10.7 ± 4.2, 0.18	28	-7.2 ± 3.3, 0.16	7	-0.79 ± 0.52, 0.28*

(b) Non-verbal

situations. Situations in which negative feelings can occur and lack enjoyment are associated with neuroticism whereas situations containing pleasant interactions are associated with extraversion, agreement, and openness [23], [24]. Additionally, stressful situations more than neutral interactions can activate the expression of neuroticism as presented by TAT [21].

2) Perceived Personality Across Scenarios and Sessions: We report results of personality inference on the scenario unseen during training in Table IV. For space, we only show results of eGeMAPS and non-verbal features as they generally outperform the speaker embedding features; speaker embedding results are in the supplemental material. We observe that the regressor performs significantly worse with all features in this cross-scenario validation despite the training and evaluation datasets containing the same participants. For most personality dimensions, the features do not explain any of the variance. Only the non-verbal feature set explains a minimal amount of variance in perceived extraversion when training on the *desk* interaction and evaluating with the *interview* interaction, indicating that certain non-verbal behaviors predictive of extraversion in the *desk* interaction generalize to the *interview* interaction. However, the overall lack of generalizability in inferring personality emphasizes that the relationship between speech and personality varies across situations.

The results of personality inference across sessions of the same conversation scenario are in Table V. For eGeMAPS features, this evaluation setup outperforms the evaluation in which both the training and evaluation data come from the same

TABLE V

(RQ3) REGRESSION RESULTS FOR PREDICTING PERSONALITY DIMENSIONS WHEN TRAINING ON ONE SESSION AND EVALUATING ON THE OTHER SESSION OF THE SAME SCENARIO. ROWS IN BOLD HIGHLIGHT THE BEST PERFORMANCE FOR THE CORRESPONDING PERSONALITY DIMENSION ACROSS BOTH FEATURE SETS. $^{\dagger}p < 0.01$; $^{*}p < 0.05$.

Scenario	Train Session	Eval. Session	Extraversion		Agreement		Conscientiousness		Neuroticism		Openness	
			Dims	R^2, r	Dims	R^2, r	Dims	R^2, r	Dims	R^2, r	Dims	R^2, r
Interview	1	2	88	0.13 ± 0.29†, 0.45†	88	0.29 ± 0.18†, 0.54†	**88**	**0.26 ± 0.14†, 0.56†**	88	-0.02 ± 0.15*, 0.26†	88	0.09 ± 0.17*, 0.36†
			31	0.11 ± 0.31†, 0.44†	**43**	**0.32 ± 0.17†, 0.57†**	28	0.20 ± 0.14†, 0.51†	6	-0.13 ± 0.10, 0.05	32	0.02 ± 0.20, 0.32†
	2	1	88	0.24 ± 0.11†, 0.54†	88	0.19 ± 0.10†, 0.47†	88	0.14 ± 0.10†, 0.40†	88	-0.04 ± 0.07, 0.17*	88	0.12 ± 0.08*, 0.36†
			27	0.24 ± 0.13†, 0.53†	52	0.21 ± 0.09†, 0.47†	39	0.14 ± 0.10†, 0.41†	15	-0.06 ± 0.09, 0.18*	**18**	**0.18 ± 0.06*, 0.45†**
Desk	1	2	88	-0.01 ± 0.11, 0.18*	88	-0.21 ± 0.37, 0.46	88	-0.09 ± 0.14, 0.06	88	0.07 ± 0.26†, 0.38†	88	-0.19 ± 0.16, -0.05
			5	-0.11 ± 0.18, 0.09	24	-0.21 ± 0.37, 0.48†	2	-0.19 ± 0.22, 0.05	21	-0.12 ± 0.35*, 0.29*	11	-0.37 ± 0.21, -0.09
	2	1	88	-0.09 ± 0.21, 0.14	88	-0.31 ± 0.36, 0.43	88	-0.08 ± 0.14, 0.08	88	0.02 ± 0.13*, 0.27†	88	-0.08 ± 0.11, 0.10
			20	-0.26 ± 0.25, 0.09	26	-0.44 ± 0.39, 0.42†	0	-	**13**	**0.14 ± 0.11†, 0.42†**	1	-0.32 ± 0.28, 0.06

(a) eGeMAPS

Scenario	Train Session	Eval. Session	Extraversion		Agreement		Conscientiousness		Neuroticism		Openness	
			Dims	R^2, r	Dims	R^2, r	Dims	R^2, r	Dims	R^2, r	Dims	R^2, r
Interview	1	2	75	0.22 ± 0.21†, 0.52†	75	0.26 ± 0.15†, 0.52†	75	0.16 ± 0.12†, 0.44†	75	-0.18 ± 0.23*, 0.03	75	0.15 ± 0.17†, 0.41†
			32	0.25 ± 0.22†, 0.55†	35	0.24 ± 0.16†, 0.50†	28	0.14 ± 0.14†, 0.41†	6	-0.26 ± 0.24, 0.02	36	0.06 ± 0.20, 0.35†
	2	1	**75**	**0.28 ± 0.11†, 0.58†**	75	0.28 ± 0.13†, 0.56†	75	0.10 ± 0.17†, 0.36†	75	-0.12 ± 0.11, 0.03*	75	0.09 ± 0.13*, 0.35†
			35	0.27 ± 0.12†, 0.57†	23	0.25 ± 0.17†, 0.53†	15	0.08 ± 0.19†, 0.36†	2	-0.17 ± 0.17, 0.07	18	0.12 ± 0.15†, 0.37†
Desk	1	2	75	-0.07 ± 0.14, 0.08	75	-0.03 ± 0.21†, 0.36†	75	-0.11 ± 0.18, 0.09	75	-0.15 ± 0.28†, 0.33†	75	-0.23 ± 0.21, 0.08
			13	-0.10 ± 0.21, 0.10	30	-0.12 ± 0.24, 0.33†	3	-0.21 ± 0.22, 0.02	28	-0.27 ± 0.28, 0.29*	7	-0.12 ± 0.17, 0.07
	2	1	75	-0.05 ± 0.18, 0.31†	75	-0.20 ± 0.22, 0.41†	75	-0.08 ± 0.17, 0.08	75	0.05 ± 0.11†, 0.31†	75	-0.17 ± 0.18, 0.21*
			10	-0.34 ± 0.33, 0.25*	16	-0.03 ± 0.18, 0.45†	1	-0.31 ± 0.24, 0.04	13	-0.03 ± 0.14, 0.17	2	-0.24 ± 0.25, 0.11

(b) Non-verbal features

session and scenario (section V-C1). This regression explains a maximum of 32% of the variance in the agreement dimension, with the performance improvement perhaps due to using the entire conversation dataset for training rather than nine of ten folds. On the other hand, regression performance improves only in the *interview* interaction with speaker embeddings and does not improve with non-verbal features.

We still observe that the *desk* interaction is more predictive of neuroticism. These results combined with the results from the cross-scenario evaluation highlight that, within a type of scenario, there are specific and relatively consistent relationships between speech features and perceived personality dimensions. However, the relationships do not generalize well across different conversation scenarios. These results showcase the importance of developing affective computing systems that are adaptable to varying contexts, as a system developed for one context may not perform well in a different context.

VI. CONCLUSION

We investigated the relationship between perceived personality and conversational speech using the UbImpressed dataset, which contains audio of the same participants in two distint conversation scenarios. Our experiments showed that perceived personality differs significantly for the same participants across a neutral and stressful interaction. We found that features related to loudness, equivalent sound level, and spectral flux are correlated with perceived extraversion, agreement, conscientiousness, and openness in the neutral scenario (*interview interaction*) and with perceived neuroticism in the stressful scenario (*desk interaction*). Through a regression analysis, we demonstrated that eGeMAPS features explain up to 32% of variance in perceived agreement, and non-verbal features explain up to 37% variance in extraversion. Furthermore, the *desk* scenario is more predictive of neuroticism while the *interview* scenario is more predictive of the remaining personality dimensions, which aligns with existing psychological research. Our consistent results across two sessions of neutral and stressful interactions emphasize the relevance of building and evaluating APP systems relative to the context in which behaviors are observed. Furthermore, our results stress the importance of context-aware affective computing systems, which has thus far been understudied.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support of the Swiss Federal Commission for Scholarships for funding this research through the Swiss Government Excellence Scholarship, grant 2024.0241. We also thank Prof. Edison Thomaz (University of Texas at Austin, USA) for feedback on the manuscript draft.

REFERENCES

[1] S. Andrist, B. Mutlu, and A. Tapus, "Look like me: Matching robot personality via gaze to increase motivation," in *CHI '15*. New York, NY, USA: ACM, 2015, p. 3603–3612.

[2] M. Braun, A. Mainz, R. Chadowitz, B. Pfleging, and F. Alt, "At your service: Designing voice assistant personalities to improve automotive user interfaces," in *CHI '19*, 2019, p. 1–11.

[3] E. C. Snyder, S. Mendu, S. S. Sundar, and S. Abdullah, "Busting the one-voice-fits-all myth: Effects of similarity and customization of voice-assistant personality," *International Journal of Human-Computer Studies*, vol. 180, p. 103126, 2023.

[4] R. R. McCrae and O. P. John, "An introduction to the five-factor model and its applications," *Journal of Personality*, 1992.

[5] G. Mohammadi and A. Vinciarelli, "Automatic personality perception: Prediction of trait attribution based on prosodic features extended abstract," in *ACII 2015*, 2015, pp. 484–490.

[6] L. Chen, R. Zhao, C. W. Leong, B. Lehman, G. Feng, and M. E. Hoque, "Automated video interview judgment on a large-sized corpus collected online," in *ACII 2017*, 2017, pp. 504–509.

[7] A. Kachur, E. Osin, D. Davydov, K. Shutilov, and A. Novokshonov, "Assessing the big five personality traits using real-life static facial images," *Scientific Reports*, vol. 10, 2020.

[8] R. Subramanian, Y. Yan, J. Staiano, O. Lanz, and N. Sebe, "On the relationship between head pose, social attention and personality prediction for unstructured and dynamic group interactions," in *ICMI 2013*, 2013, p. 3–10.

[9] R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, and N. Sebe, "Ascertain: Emotion and personality recognition using commercial sensors," *IEEE Trans. on Affective Computing*, vol. 9, no. 2, pp. 147–160, 2018.

[10] A. Vinciarelli and G. Mohammadi, "A survey of personality computing," *IEEE Trans. on Affective Computing*, vol. 5, no. 3, pp. 273–291, 2014.

[11] B. Schuller, S. Steidl, A. Batliner, E. Noeth, A. Vinciarelli, F. Burkhardt, R. van Son, F. Weninger, F. Eyben, T. Bocklet, G. Mohammadi, and B. Weiss, "The INTERSPEECH 2012 speaker trait challenge," in *INTERSPEECH 2012*, 2012.

[12] W. Mischel, *Personality and Assessment*. John Wiley & Sons Inc., 1968.

[13] N. S. Endler and D. Magnusson, "Toward an interactional psychology of personality." *Psychological bulletin*, vol. 83, pp. 956–74, 1976.

[14] W. Mischel and P. Peak, "Some facets of consistency: Replies to Epstein, Funder, and Bem," *Psychological Review*, vol. 90, pp. 394–402, 1983.

[15] S. Muralidhar, L. S. Nguyen, D. Frauendorfer, J.-M. Odobez, M. Schmid Mast, and D. Gatica-Perez, "Training on the job: behavioral analysis of job interviews in hospitality," in *ICMI '16*, 2016, p. 84–91.

[16] S. Muralidhar, M. S. Mast, and D. Gatica-Perez, "A tale of two interactions: Inferring performance in hospitality encounters from cross-situation social sensing," *IMWUT*, vol. 2, no. 3, Sep. 2018.

[17] S. Muralidhar, R. Siegfried, J.-M. Odobez, and D. Gatica-Perez, "Facing employers and customers: What do gaze and expressions tell about soft skills?" in *Proc. of the 17th International Conference on Mobile and Ubiquitous Multimedia*, ser. MUM '18, 2018, p. 121–126.

[18] W. Fleeson and E. Noftle, "The end of the person–situation debate: An emerging synthesis in the answer to the consistency question," *Social and Personality Psychology Compass*, 2008.

[19] W. Fleeson, "Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability," *Current Directions in Psychological Science*, 2004.

[20] W. Mischel, E. B. Ebbesen, and A. R. Zeiss, "Selective attention to the self: Situational and dispositional determinants." *Journal of Personality and Social Psychology*, vol. 27, no. 1, p. 129, 1973.

[21] R. Tett and D. Burnett, "A personality trait-based interactionist model of job performance," *Journal of Applied Psychology*, 2003.

[22] T. A. Judge and C. P. Zapata, "The person–situation debate revisited: Effect of situation strength and trait activation on the validity of the big five personality traits in predicting job performance," *Academy of Management Journal*, vol. 58, no. 4, pp. 1149–1179, 2015.

[23] J. Rauthmann, D. Gallardo-Pujol, E. Guillaume, E. Todd, C. Nave, R. Sherman, M. Ziegler, A. Jones, and D. Funder, "The situational eight DIAMONDS: A taxonomy of major dimensions of situation characteristics," *Journal of Personality and Social Psychology*, vol. 107, pp. 677–718, 2014.

[24] S. Parrigon, S. E. Woo, L. Tay, and T. Wang, "CAPTION-ing the situation: A lexically-derived taxonomy of psychological situation characteristics," *Journal of personality and social psychology*, vol. 112, 2017.

[25] J. A. Shaffer and B. E. Postlethwaite, "The validity of conscientiousness for predicting job performance: A meta-analytic test of two hypotheses," *International Journal of Selection and Assessment*, vol. 21, no. 2, pp. 183–199, 2013.

[26] V. Pace and M. Brannick, "Improving prediction of work performance through frame-of-reference consistency: Empirical evidence using openness to experience," *International Journal of Selection and Assessment*, vol. 18, pp. 230 – 235, 06 2010.

[27] N. Bowling and G. Burns, "A comparison of work-specific and general personality measures as predictors of work and non-work criteria," *Personality and Individual differences*, vol. 49, pp. 95–101, 07 2010.

[28] J. Pohjalainen, S. Kadioglu, and O. Räsänen, "Feature selection for speaker traits," in *Proc. Interspeech*, 09 2012, pp. 270–273.

[29] A. Ivanov and X. Chen, "Modulation spectrum analysis for speaker personality trait recognition," in *INTERSPEECH 2012*, 2012, pp. 278–281.

[30] M. Tayarani, A. Esposito, and A. Vinciarelli, "What an "ehm" leaks about you: Mapping fillers into personality traits with quantum evolutionary feature selection algorithms," *IEEE Trans. on Affective Computing*, vol. 13, no. 1, pp. 108–121, 2022.

[31] M.-A. Carbonneau, E. Granger, Y. Attabi, and G. Gagnon, "Feature learning from spectrograms for assessment of personality traits," *IEEE Trans. on Affective Computing*, vol. 11, no. 1, pp. 25–31, 2020.

[32] H. Le, S. Li, C. O. Mawalim, H.-H. Huang, C. W. Leong, and S. Okada, "Investigating the effect of linguistic features on personality and job performance predictions," in *Social Computing and Social Media*. Berlin, Heidelberg: Springer-Verlag, 2023, p. 370–383.

[33] R. Barchi, L. Pepino, M. Gauder, L. Estienne, P. Riera, and L. Ferrer, "Apparent personality prediction from speech using expert features and wav2vec 2.0," in *Workshop on Speech, Music and Mind*, 2023, pp. 21–25.

[34] J.-I. Biel and D. Gatica-Perez, "The YouTube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs," *IEEE Trans. on Multimedia*, vol. 15, no. 1, pp. 41–55, 2013.

[35] ——, "Mining crowdsourced first impressions in online social video," *IEEE Trans. on Multimedia*, vol. 16, no. 7, pp. 2062–2074, 2014.

[36] A. Guidi, C. Gentili, E. Scilingo, and N. Vanello, "Analysis of speech features and personality traits," *Biomedical Signal Processing and Control*, vol. 51, pp. 1–7, 05 2019.

[37] I. Naim, M. I. Tanveer, D. Gildea, and M. E. Hoque, "Automated prediction and analysis of job interview performance: The role of what you say and how you say it," in *FG*, vol. 1, 2015, pp. 1–6.

[38] P. E. Shrout and J. L. Fleiss, "Intraclass correlations: uses in assessing rater reliability." *Psychological bulletin*, vol. 86, no. 2, p. 420, 1979.

[39] S. Muralidhar, M. Schmid Mast, and D. Gatica-Perez, "How may I help you? Behavior and impressions in hospitality service encounters," in *ICMI '17*. New York, NY, USA: ACM, 2017, p. 312–320.

[40] J. L. Hodges, "The significance probability of the smirnov two-sample test," *Arkiv för Matematik*, vol. 3, pp. 469–486, 1958.

[41] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, "Using linguistic cues for the automatic recognition of personality in conversation and text," *J. Artif. Int. Res.*, vol. 30, no. 1, p. 457–500, Nov. 2007.

[42] F. Eyben, K. Scherer, B. Schuller, J. Sundberg, E. Andre, C. Busso, L. Devillers, J. Epps, P. Laukka, S. Narayanan, and K. Truong, "The geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing," *IEEE Trans. on Affective Computing*, 2016.

[43] F. Eyben, M. Wöllmer, and B. Schuller, "openSMILE – the munich versatile and fast open-source audio feature extractor," in *Proc. of the ACM Multimedia 2010 International Conference*, 2010, pp. 1459–1462.

[44] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, and S. Khudanpur, "X-vectors: Robust dnn embeddings for speaker recognition," in *ICASSP 2018*, 2018, pp. 5329–5333.

[45] B. Desplanques, J. Thienpondt, and K. Demuynck, "ECAPA-TDNN: Emphasized channel attention, propagation and aggregation in tdnn based speaker verification," *INTERSPEECH 2020*, 2020.

[46] I. Ulgen, Z. Du, C. Busso, and B. Sisman, "Revealing emotional clusters in speaker embeddings: A contrastive learning strategy for speech emotion recognition," in *ICASSP 2024*, 04 2024, pp. 12 081–12 085.

[47] H. Bredin, R. Yin, J. M. Coria, G. Gelly, P. Korshunov, M. Lavechin, D. Fustes, H. Titeux, W. Bouaziz, and M.-P. Gill, "pyannote.audio: neural building blocks for speaker diarization," in *ICASSP 2020*, 2020.

[48] T. Özseven, M. Düğenci, A. Doruk, and H. I. Kahraman, "Voice traces of anxiety: acoustic parameters affected by anxiety disorder," *Archives of Acoustics*, pp. 625–636, 2018.

[49] E. Gilboa-Schechtman and I. Shachar-Lavie, "More than a face: a unified theoretical perspective on nonverbal social cue processing in social anxiety," *Frontiers in Human Neuroscience*, vol. 7, 2013.

[50] S. Hofmann, A. Gerlach, A. Wender, and W. Roth, "Speech disturbances and gaze behavior during public speaking in subtypes of social phobia," *J Anxiety Disord*, vol. 11, pp. 573–585, 1997.

[51] M. Lewin, D. McNeil, and L. J.M., "Enduring without avoiding: Pauses and verbal dysfluencies in public speaking fear," *J Psychopathol Behav Assess*, vol. 18, pp. 387–402, 1996.

[52] Q. Luo, Y. Di, and T. Zhu, "Predictive modeling of neuroticism in depressed and non-depressed cohorts using voice features," *Journal of Affective Disorders*, vol. 352, pp. 395–402, 2024.

[53] C. Mawalim, S. Okada, Y. Nakano, and M. Unoki, "Personality trait estimation in group discussions using multimodal analysis and speaker embedding," *Journal on Multimodal User Interfaces*, 2023.

Alice Zhang is a Ph.D. student at the University of Texas at Austin and a visiting Ph.D. student at Idiap Research Institute. Her research interests include acoustic sensing and wearable technologies to support and analyze social interactions.

Skanda Muralidhar is a researcher at Idiap Research Institute, Switzerland, where he researches and develops AI-powered technologies for sign language learning. He holds a Ph.D. in Electrical Engineering from EPFL, with expertise in multimodal signal processing, human emotion recognition, and machine learning. His current work advances applied AI for education, healthcare, and human-centered innovation.

Daniel Gatica-Perez directs the Social Computing Group at Idiap Research Institute and is a professor at EPFL, Switzerland. His research integrates human-centered and participatory methods with mobile, social, and AI technologies to support individuals and communities. He also works with organizations on social innovation projects.

Mathew Magimai-Doss received his Ph.D. in Electrical Engineering from EPFL, Switzerland in 2005. He is now a Senior Researcher at Idiap Research Institute, Switzerland. His research interests lie in signal processing, statistical pattern recognition, artificial neural networks and computational linguistics with applications to speech, audio, and multimodal signal processing.