

Toward Multimodal Modeling of Emotional Expressiveness

Victoria Lin*
Carnegie Mellon University
vlin2@andrew.cmu.edu

Michael A. Sayette
University of Pittsburgh
sayette@pitt.edu

Jeffrey M. Girard*
Carnegie Mellon University
jmgirard@cmu.edu

Louis-Philippe Morency
Carnegie Mellon University
morency@cs.cmu.edu

ABSTRACT

Emotional expressiveness captures the extent to which a person tends to outwardly display their emotions through behavior. Due to the close relationship between emotional expressiveness and behavioral health, as well as the crucial role that it plays in social interaction, the ability to automatically predict emotional expressiveness stands to spur advances in science, medicine, and industry. In this paper, we explore three related research questions. First, how well can emotional expressiveness be predicted from visual, linguistic, and multimodal behavioral signals? Second, how important is each behavioral modality to the prediction of emotional expressiveness? Third, which behavioral signals are reliably related to emotional expressiveness? To answer these questions, we add highly reliable transcripts and human ratings of perceived emotional expressiveness to an existing video database and use this data to train, validate, and test predictive models. Our best model shows promising predictive performance on this dataset ($RMSE = 0.65$, $R^2 = 0.45$, $r = 0.74$). Multimodal models tend to perform best overall, and models trained on the linguistic modality tend to outperform models trained on the visual modality. Finally, examination of our interpretable models' coefficients reveals a number of visual and linguistic behavioral signals—such as facial action unit intensity, overall word count, and use of words related to social processes—that reliably predict emotional expressiveness.

CCS CONCEPTS

• Applied computing → Psychology.

KEYWORDS

affective computing, behavior understanding, emotion expression

Victoria Lin, Jeffrey M. Girard, Michael A. Sayette, and Louis-Philippe Morency. 2020. Toward Multimodal Modeling of Emotional Expressiveness. In *ICMI '20: ACM International Conference on Multimodal Interaction, October 25–29, 2020, Utrecht, the Netherlands. ACM, New York, NY, USA, 10 pages*. <https://doi.org/10.1145/1122445.1122456>

*Both authors contributed equally to this research.

1 INTRODUCTION

Emotional expressiveness is a psychological attribute that captures the degree to which a person tends to display their emotions through behavior [25, 36]. A highly expressive person tends to exhibit what they feel through their facial movements, vocalizations, and other gestures, whereas a less expressive person tends to hide or hold their feelings in. Emotional expressiveness is related to but distinct from other affect-related traits. It is much narrower in its scope (i.e., more specific in terms of the patterns of behavior it explains) than broad personality traits such as extraversion and neuroticism [40], and it differs from emotionality [58] in that it describes the frequency and intensity of emotional *displays* rather than the frequency and intensity of emotional *experiences*.

Greater scientific understanding of emotional expressiveness is critical for many areas of basic and applied research. For example, correctly inferring a person's mental state from their behavior (e.g., how much they are enjoying a comedy) likely requires careful consideration of how expressive that person is generally: a small chuckle from a highly expressive person may mean that they are unimpressed, whereas that same chuckle from a normally tight-lipped person may mean the opposite. Similarly, when synthesizing expressive behavior for a robot or virtual agent, it is important to consider how much emotion it should display over the course of an interaction. For example, there may be some settings where a more expressive robot or agent would increase user engagement and enjoyment [28] and others where such displays would be inappropriate or distracting. Users may also prefer a robot or agent that matches his or her level of expressiveness, which would require the ability to detect user expressiveness and then modulate the synthesis of behavior accordingly. Finally, emotional expressiveness plays an important role in healthcare, where persistent abnormalities in the display of emotion (e.g., reduced or inappropriate emotional expressiveness) serve as observable signs of various medical conditions (e.g., depression, mania, schizophrenia, and brain injury) [1]. The ability to automatically detect such abnormalities could enhance the objectivity and efficiency of clinical assessment.

Furthermore, in many of these applications, interpretability is paramount. Understanding which behavioral signals are reliably related to emotional expressiveness would enable virtual agents and robots to more effectively simulate expressive behavior and would aid in the training of clinicians. Understanding which behavioral signals a model is using to predict emotional expressiveness would help to validate these models and would allow clinicians to

discount their predictions when appropriate (e.g., when an important behavioral signal observed in a given patient is known to be caused by something other than disease, such as a drug side-effect).

In an ideal world, emotional expressiveness would be objectively measured as the *true* frequency and intensity of emotional displays. However, the affective sciences currently lack a single, agreed-upon way of delineating and representing emotional displays [4, 10], so this objective ground truth is currently out of reach. Instead, we focus on third-party observers’ *perceptual ratings* of emotional expressiveness. This approach has the potential to introduce subjectivity and bias, but we can quantify this subjectivity through inter-rater reliability analyses and mitigate it through the aggregation of ratings [38, 51]. Operationalizing emotional expressiveness through observer ratings also has the benefit of paralleling the standard approach used in medicine, where clinicians observe patients’ behavior and rate their emotional expressiveness [1].

Motivated by the potential applications described previously, the current study explores the following research questions:

- (1) How well can we predict emotional expressiveness from multimodal representations of participants’ behavior?
- (2) How much do the different behavioral modalities contribute to the prediction of emotional expressiveness?
- (3) Which behavioral signals are reliably related to emotional expressiveness? What is the nature of these relationships?

To begin answering these questions, the current study makes three primary contributions to the literature. First, we add highly reliable transcripts and perceptual ratings of emotional expressiveness to an existing video database. Second, we train algorithms to predict emotional expressiveness from visual, linguistic, and multimodal representations of participants’ behavior and compare their performance. Third, we explore what our best-performing algorithm learned about the interpretable relationships between behavioral signals and perceptual ratings of emotional expressiveness.

2 RELATED WORK

Although computational work on emotional expressiveness has been rare, interdisciplinary researchers have extensively studied the adjacent topics of personality perception and emotion recognition. Both topics resemble the task of computationally characterizing emotional expressiveness, as they involve inferring psychological traits and states, respectively, from observed behavior. Therefore, we look to research in these areas to inform our own approach to predicting and characterizing emotional expressiveness. In this section, we describe the current state of these areas of study.

2.1 Personality Perception

In personality computing, the prediction of personality traits as rated by external observers is referred to as *automatic personality perception* [64]. Studies in automatic personality computing typically classify whether an individual is above or below the median on a trait (although a more rigorous approach is to predict their dimensional scores [12, 65]). Previous models used acoustic [49, 54], linguistic [61], and visual [5] data, often in multimodal combinations. Traditional machine learning algorithms have been most common, especially support vector machines (SVMs) [6, 45, 56], although some interpretable approaches have been used [44].

Recently, deep learning approaches have played a larger role in personality computing. For example, in the ChaLearn LAP challenge [50], four models achieved competitive performance out of eight participating models. All four of these top-performing models used convolutional neural networks (CNNs) as feature extractors; to predict over these features, three models used deep learning regressors, and one used support vector regression (SVR).

2.2 Emotion Recognition

Generally, efforts in emotion recognition involve either the prediction of discrete emotion categories or continuous affective dimensions from short segments of behavior [26]. Notably, emotion recognition differs from both automatic personality perception and the emotional expressiveness prediction task in this paper in that it involves computing a momentary state rather than a longer-lasting trait. Nonetheless, emotion recognition shares many methodological similarities with automatic personality perception.

Emotion recognition models have been trained on acoustic, linguistic, visual, and multimodal data [67]. The most common approaches are machine learning classifiers including SVMs [31, 53, 63], deep belief networks [7, 59], and ensemble methods like AdaBoost [39, 48, 66]. Recently, partially in response to emotion recognition challenges from video like EmotiW [13–15] and AVEC [62], multimodal deep learning models—often comprised of CNN feature extractors over either a deep learning classifier [60] or a simpler method like SVM [55] or logistic regression [16]—have assumed a prominent role in emotion recognition tasks, particularly in the classification of basic emotion categories. Extensions of this type of architecture to multitask learning and cascade networks, as well as experiments with generative adversarial networks and variational autoencoders, produce the current state-of-the-art results [37].

2.3 Emotional Expressiveness

Prior work in psychology focused on refining the definition and measurement of emotional expressiveness as a trait. Several questionnaires were developed to measure emotional expressiveness via self-report and were used to examine its relationship with other affective and personality attributes [21, 25, 33, 36]. These studies found that a general disposition to outwardly display emotions could be readily found and measured via self-report. This disposition was found to be most closely aligned with the personality trait of extraversion (i.e., sociability, assertiveness, and activity), although it was also related to neuroticism (i.e., anxiety, hostility, and impulsiveness) and agreeableness (i.e., trust, compliance, and modesty) in different contexts. Specifically, extraversion appears to be related to increased expression of both positive and negative emotion, whereas neuroticism appears to favor the expression of negative emotion, and agreeableness appears to favor the expression of positive emotion [25]. It is important to note, however, that these relationships were all modest in magnitude (e.g., correlations were around 0.2 to 0.4 with extraversion, -0.2 to 0.3 with neuroticism, and 0.0 to 0.2 with agreeableness), which supports the idea that emotional expressiveness is related to but still distinct from these broader personality traits.

Human observers have also been used to measure how emotionally expressive a person appears to be within a given interaction



Figure 1: Example frames from the GFT database

[36, 57]. However, these studies did not explore which specific behavioral signals the observers used to make their ratings. It is thus unclear which aspects of communicative behavior are most expressive. Recently, we sought to address this gap in the literature by automatically predicting facial expressiveness and quantifying its behavioral correlates on a moment-to-moment basis using the Elastic Net algorithm [38]. This work informs our choice of methods in the current paper but differs in a number of important ways. First, whereas the previous study examined expressiveness using only visual representations of behavior, the current study also examines linguistic and multimodal representations—both of which have been shown to be crucial in studies on personality computing and emotion recognition—in addition to visual representations. Second, whereas the previous study examined expressiveness on a momentary time-scale, the current study examines holistic patterns of expressiveness over a longer time period to better capture emotional expressiveness as an enduring pattern of behavior. Finally, whereas the previous study examined induced responses to emotional stimuli, the current study examines expressive behavior in a more naturalistic and unstructured social interaction setting.

3 DATA

The videos examined in this work came from the Sayette Group Formation Task (GFT) spontaneous facial expression database [24]. This database includes videos of 96 participants (42% Female, 85% White) observed during a psychology experiment [52]. In this experiment, a group of three participants would meet for the first time and get to know one another over the course of 36 min; this time was unstructured, and groups could discuss whatever topics they wished. The GFT database includes 1 min videos of each group, starting an average of 5.6 min into the interaction. Separate wall-mounted cameras faced each participant and recorded a 720×480 pixel video of their head and shoulders at a rate of 29.97 fps. Example frames from these videos can be found in Figure 1.

Audio during the experiment was recorded from a single centrally located microphone.¹ In order to measure language-related behaviors, we manually transcribed text from the audio recordings. The speech of all 96 participants was transcribed by the first author and, for the purposes of assessing transcription reliability, the speech from a subset of 30 participants was also transcribed by the

¹The audio recordings and personality scores were not released as part of the official GFT database distribution [24]. We were granted special access for this study.

second author. Word-level agreement was high (95.1%) on this subset of participants, which provides evidence that the transcriptions are trustworthy overall.

Self-reported personality traits were measured using the NEO-FFI [9], a 60-item inventory yielding scale scores on the “Big Five” personality traits: neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness traits.¹ This inventory was completed by participants prior to the interaction.

4 MEASURING EXPRESSIVENESS

In order to measure how emotionally expressive each participant in the GFT dataset was, we paid human annotators to watch and rate each video. To mitigate the influence of annotator subjectivity, each video was rated by multiple annotators, and ratings were averaged per video [51]. Moreover, to increase the reliability and validity of these ratings, annotators completed multiple items (i.e., questions) measuring different aspects of emotional expressiveness, which we combined into a single score per video using latent variable modeling [34]. Although they take longer to complete, multiple-item scales are generally preferred to single-item scales because the latter tend to be quite noisy (i.e., lack reliability), struggle to capture the complexity and nuance of many attributes (i.e., lack validity), and can lead to incorrect analytical conclusions [11]. In this section, we describe the process of collecting these perceptual ratings and conducting the latent variable analysis.

4.1 Perceptual Ratings

Human annotators were recruited through the *Prolific.co* online research platform. In order to better assess annotator quality and consistency, we wanted each annotator to rate multiple videos. However, requiring each annotator to rate all the videos might lead to fatigue or dropout. To balance these two concerns, we separated the 96 participant videos into 6 sets of 16 videos and assigned a different group of annotators to rate each set. To minimize the influence of cultural and sensory differences, annotators were required to have USA nationality, English language fluency, and normal (or corrected to normal) vision and hearing. To ensure high-quality ratings, annotators were required to have a prior approval rate of 95% or higher, and annotators were replaced if they failed one or more attention check questions during our experiment. For each set of videos, we recruited eight crowdworkers to serve as raters. After three annotators were replaced for failing attention checks, the final sample of 48 annotators was 63% Female, 60% White, and ranged in age from 19 to 60 years old ($M = 30.2$, $SD = 10.3$).

Using the *formr* online survey platform [2], annotators viewed instructions and then rated each video in their assigned set. Annotators were instructed to watch each video (from start to finish at normal speed and with the volume on) and then carefully answer a series of 12 questions about how they perceived the person in the video. In this study, we focused on a subset of 4 of these questions that captured different aspects of emotional expressiveness.

Although previous work described in subsection 2.3 developed methods for measuring emotional expressiveness, many of these self-report items were difficult to adapt to observational research (e.g., “The way I feel is different from how others think I feel”) and previous observational studies used single-item measures (e.g.,

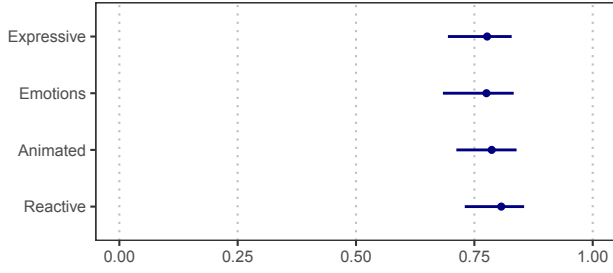


Figure 2: Inter-annotator reliability as measured by intra-class correlation coefficients (with 95% confidence intervals)

rating videos on “expressiveness”), which we previously noted can attenuate reliability and validity. As such, we decided to design new questions to be used in a multiple-item scale. We chose four items to balance the competing needs for comprehensiveness (which requires more items) and efficiency (which requires fewer).

- (1) How **expressive** was the person in this video (use your own understanding of what it means to be expressive)?
- (2) How much did the person in this video show their **emotions** (through their words and nonverbal behavior)?
- (3) How **animated** (lively, energetic, or active) was the person in this video?
- (4) How much did the person in this video **react** to the other people (through their words and nonverbal behavior)?

The first question represented the single-item measures used in previous observational studies, and the second question captured the theoretical definition of emotional expressiveness. The third and fourth questions were more exploratory and were our attempts to represent other observable aspects of emotional expressiveness. Specifically, the third question attempted to capture the idea that emotions are often strongly expressed through movement and activity, and the fourth question attempted to capture the idea that emotions (and therefore emotional expressions) are reactions to environmental (and often social) stimuli. All questions were carefully worded to generalize across all emotions and behavioral modalities. Each question was answered using a rating scale with five categorical options ranging from 0 (*not at all*) to 4 (*extremely*).

The inter-annotator reliability of the perceptual ratings was calculated per question using the *agreement* software package [23]. Intraclass correlation coefficients (ICCs) under model 2A [27] were calculated to estimate the reliability of the average of all eight annotators’ scores. ICC values above 0.50, 0.75, and 0.90 are generally considered to be evidence of moderate, good, and excellent reliability, respectively [35]. Note that these ICCs are more conservative than Pearson correlation coefficients [41]. As shown in Figure 2, reliability estimates were all in the good range, and confidence intervals (CIs) did not extend below the moderate range.

4.2 Latent Variable Modeling

After being averaged across raters, scores on the four expressiveness-related questions were highly inter-correlated (all $r > 0.83$), which supports the feasibility of combining these scores into a single lower-dimensional representation. The simplest way to combine

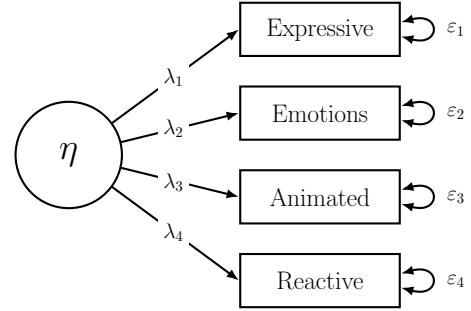


Figure 3: Path diagram of the CFA model (ovals are latent variables, rectangles are manifest/observed variables, λ are factor loadings, and ϵ are residual variances)

the scores would have been to sum or average them, but to do so would assume that all questions are equally important and equally well-measured; these assumptions are unlikely to be met in practice and would negatively impact the validity and reliability of the aggregate if violated [42]. Thus, to avoid these assumptions, we used confirmatory factor analysis (CFA) [34] to estimate a latent variable η that explains the variance shared among the questions (Figure 3). In this model, factor loadings represent how much of each question was composed of shared variance (i.e., how strong the relationship was between the question and the latent variable), and residual variances represent how much each question was composed of non-shared variance (i.e., unique variance and measurement error). Because these model parameters are freely estimated (i.e., informed by the data), the latent variable η will be a better representation of the shared variance than a sum or average would be [42].

The CFA model was estimated in a Bayesian framework using Markov Chain Monte Carlo in the *blavaan* software package [43]. All latent and manifest variables were standardized to zero mean and unit variance. Normal priors ($\mu = 0$, $\sigma = 1$) were used for the intercept and factor loading parameters, and gamma priors ($\alpha = 1$, $\beta = 1$) were used for the precision (i.e., residual standard deviation) parameters. Four Markov chains were estimated, each with 1000 warmup and 1000 inference iterations.

We evaluated the fit of the CFA model using Bayesian versions of two common fit indexes that are known to not be biased when calculated in small samples [22]: the gamma-hat fit index ($\hat{\Gamma}$), which is based on the noncentral chi-square distribution, and the comparative fit index (CFI), which compares the proposed model to both a baseline model (in which all covariances are set to zero) and a saturated model (in which all covariances are freely estimated). Both of these indexes range from 0 to 1 and are commonly interpreted as being indicative of good model fit when above 0.950 [30].

The fit of the CFA model was good according to both fit indexes ($\hat{\Gamma} = 0.981$, CFI = 0.993). Table 1 provides a summary of the model’s parameter estimates. All four questions had strongly positive loadings on the latent variable, which shows that they shared a great deal of variance (i.e., tended to vary together) and supports our argument that they measured a single underlying attribute [20].

To further validate our latent variable as a measure of emotional expressiveness (as opposed to some other quantity), we extracted

Table 1: Results from the Bayesian CFA model

Parameter		Estimate	95% CI
λ_1	Expressive Factor Loading	0.97	[0.84, 1.14]
λ_2	Emotions Factor Loading	0.95	[0.81, 1.11]
λ_3	Animated Factor Loading	0.96	[0.83, 1.13]
λ_4	Reactive Factor Loading	0.87	[0.72, 1.05]
ϵ_1	Expressive Residual Variance	0.07	[0.04, 0.10]
ϵ_2	Emotions Residual Variance	0.11	[0.07, 0.16]
ϵ_3	Animated Residual Variance	0.08	[0.05, 0.12]
ϵ_4	Reactive Residual Variance	0.24	[0.18, 0.33]

factor scores (i.e., estimates of each participant’s standing on the latent variable η) and assessed their pattern of correlation with participants’ self-reported personality traits. If our latent variable is a good estimate of emotional expressiveness, this pattern of correlation should match expectations based on theory and previous work [20]. Based on the literature reviewed in subsection 2.3, we expected the factor scores to be moderately positively correlated with extraversion, weakly positively correlated with agreeableness, and weakly positively or negatively correlated with neuroticism depending on the negativity versus positivity of the context.

The factor scores were positively correlated with extraversion (0.26) and agreeableness (0.28) and slightly negatively correlated with neuroticism (−0.07). These results largely matched expectations, although the correlation with agreeableness was a bit larger than expected. These differences may be due in part to the context of the GFT study, wherein participants interacted with strangers (e.g., agreeable participants may be especially expressive in this context due to higher interpersonal trust and cooperation). The weak negative correlation with neuroticism suggests that this context may attenuate displays of negative emotion (e.g., to make a good first impression). Given this evidence of the latent variable’s internal consistency and external validity, we felt confident using the factor scores as labels of emotional expressiveness in our predictive modeling experiments [42].

5 MEASURING BEHAVIOR

In order to study the ways in which emotional expressiveness manifests in communicative behavior, we needed to measure participants’ behavior. To do so, we designed measures of various visual and linguistic behavioral signals from the video recordings and transcripts. We based our choice of behavioral signals on measures that had been used successfully in personality perception and emotion recognition, as described in the literature reviewed in Section 2.3. We decided not to include measures of acoustic behaviors (e.g., speech prosody) because of the high likelihood of imperfect audio source separation among the three participants in each group, particularly given the suboptimal recording quality and the frequency with which subjects spoke simultaneously. However, we do believe that acoustic behavioral signals are important to expressiveness and plan to explore them in future work after experimenting with different approaches for source separation.

5.1 Visual Behavioral Signals

For each participant’s video, we used the OpenFace 2.0 [3] toolkit to derive measures of visual behavioral signals. We began with frame-level estimates of the occurrence and intensity of action units (i.e., facial muscle movements) from the Facial Action Coding System [19], facial landmark (e.g., eye and mouth) coordinates, head pose (i.e., translation and rotation), and eye gaze angle. To reduce the effects of jitter, we averaged these measures over five-frame intervals, reducing the sampling rate from 30 Hz to 6 Hz. Furthermore, we used an affine transformation [29] to project each participant’s facial landmark coordinates onto the average of all participant faces. This normalized the faces for size, translation, and rotation and isolated landmark movement from head movement.

We then calculated the average number of action units present and the mean of all action units’ intensities across all intervals. Our decision to aggregate all action units was motivated by a desire to capture overall facial movement without requiring OpenFace to reliably distinguish individual action units, which is a difficult task (especially when generalizing to novel datasets) [8]. To measure the overall magnitude and speed of participants’ facial and head motion, we computed the average displacement (i.e., distance traveled), velocity (i.e., the derivative of displacement), and acceleration (i.e., the derivative of velocity) for each facial landmark, head pose dimension, and eye gaze angle across five-frame intervals. Because measures for the individual facial landmark points were highly correlated, we summed across all landmark points. A full list of the 20 included visual behavioral signals can be found in Table 2.

5.2 Linguistic Behavioral Signals

For each participant’s transcript, we used the LIWC2015 [47] text analysis program to derive measures of their linguistic behavior across all utterances. LIWC2015 returns many measures that represent the percentage of the words in the text that fall into different categories (e.g., personal pronouns, common verbs, or swear words) as well as the output of several algorithms trained to predict linguistic dimensions (e.g., emotional tone, authenticity, and clout).

We selected categories and dimensions to include in our analyses based on several considerations. Because we were analyzing spoken words, we excluded measures that assume written text (e.g., punctuation and words per sentence). We prioritized categories and dimensions related to affective, social, and psychological processes (e.g., emotional tone, affiliation, and cognition). We also included categories related to parts of speech and language construction (e.g., pronouns, verbs, and fillers).

In general, we chose to group linguistic behavioral signals into broader categories rather than including each individual category returned by LIWC2015 in order to reduce dimensionality and increase category reliability through aggregation. A full list of the 35 included linguistic behavioral signals can be found in Table 2.

6 PREDICTIVE MODELING

We used predictive modeling techniques to explore our three research questions: (1) How well can we predict emotional expressiveness from multimodal representations of participants’s behavior? (2) How much do the different behavioral modalities contribute to

Table 2: List of behavioral signals by modality

Visual	Linguistic	
Mean Number of Action Units	Word Count	Cognitive Processes
Mean Action Unit Intensity	Analytical Thinking	Sexual
Mean Landmark Displacement	Clout	Ingestion
Mean Landmark Velocity	Authentic	Affiliation
Mean Landmark Acceleration	Emotional Tone	Reward
Head Translation Displacement	Words > 6 Letters	Risk
Head Translation Velocity	Personal Pronouns	Past Focus
Head Translation Acceleration	Impersonal Pronouns	Present Focus
Head Pitch Displacement	Common Adverbs	Future Focus
Head Pitch Velocity	Negations	Work
Head Pitch Acceleration	Common Verbs	Leisure
Head Yaw Displacement	Common Adjectives	Home
Head Yaw Velocity	Comparisons	Money
Head Yaw Acceleration	Interrogatives	Swear words
Head Roll Displacement	Quantifiers	Assent
Head Roll Velocity	Positive Emotion	Nonfluencies
Head Roll Acceleration	Negative Emotion	Fillers
Gaze Angle Displacement	Social Processes	
Gaze Angle Velocity		
Gaze Angle Acceleration		

the prediction of emotional expressiveness? (3) Which behavioral signals are reliably related to emotional expressiveness?

6.1 Algorithms

To investigate our first two questions, we needed predictive modeling algorithms that are well understood and well performing when trained on relatively small datasets. To investigate our third question, we needed at least one algorithm that could achieve competitive performance while providing interpretable results regarding the direction and strength of the relationship between each feature and label. Based on our literature review and prior experience, we decided to include the three algorithms described below.²

Elastic Net. This algorithm uses a mixture of L1 and L2 regularization on linear regression to reduce overfitting and multicollinearity, both of which are common problems when fitting linear models to data with many features [68]. Elastic Net models are fully interpretable, with the regression coefficients providing an indication of the direction and strength of the relationship between the features and the label. Our past work on visual expressiveness [38] found that Elastic Net’s predictive performance exceeded that of several more complex models when trained on a small dataset.

Support Vector Regression (SVR). This algorithm transforms data into a higher-dimensional feature space and constructs a hyperplane that maximizes the number of points within some distance from it [17]. Because Elastic Net models linear relationships only, we selected SVR with a radial basis function (RBF) kernel as a competing model to see whether modeling nonlinearities would yield better performance (at the cost of being able to interpret the relationships between features and labels). As described in Section 2, SVR and SVM have seen widespread use in the related fields of automatic personality perception and emotion recognition.

²Our goal in comparing multiple algorithms was not to determine which is best in general but rather to find one that performed well in this application. Proposing new algorithms is an interesting direction for future research but is beyond the scope of the current paper. Here, we focus on introducing a new topic area to the community and using existing algorithms to deepen scientific understanding of it.

Multi-Layer Perceptron (MLP). Neural network models currently set the standard of performance for many relevant prediction tasks (as described in Section 2) due to their ability to capture complex nonlinear relationships (at the cost of interpretability, like SVR). Due to the relatively small size of our dataset and the absence of a temporal component, we constrained the complexity of our model to a simple MLP with one or two hidden layers.

6.2 Experimental Setup

To explore our second research question about the relative importance of each modality, we fit three models per algorithm: one model using only visual behavioral signals as features, one model using only linguistic behavioral signals as features, and one multimodal model using both sets of behavioral signals as features.

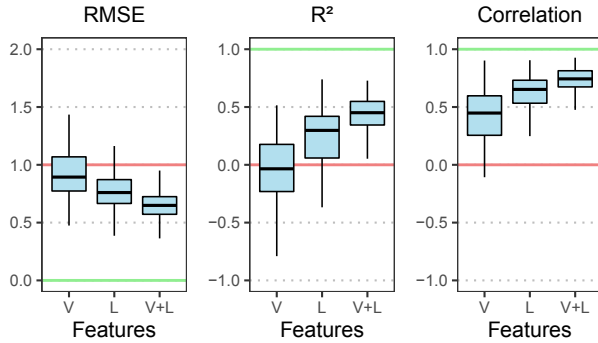
We performed 7-fold validation within 8-fold testing and repeated this process 20 times with different splits of the data. The folds were stratified to keep the proportion of labels in each quartile similar in each fold, and all three members of a given group were always assigned to the same fold (in order to prevent group-level dependencies from influencing performance estimates). Within each repetition, the folds were identical across algorithms and modalities to ensure that results were directly comparable. Root mean square error (RMSE) was used as the validation error measure.

We used the *sklearn* implementations of these algorithms in our experiments and tuned hyperparameters via grid search. For Elastic Net, we tuned over $\alpha \in \{0.01, 0.05, 0.1, 0.5, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5\}$ and $\lambda \in \{0.0, 0.1, 0.5, 0.7, 0.9, 0.95, 0.99, 1.0\}$, where α is the penalty term weight and λ is the mixing parameter for L1 and L2 regularization. When $\lambda = 0.0$, Elastic Net becomes ridge regression, and when $\lambda = 1.0$, it becomes lasso regression [68]. For SVR, we tuned over $C \in \{2^{-5}, 2^{-4}, \dots, 2^{14}, 2^{15}\}$ and $\gamma \in \{2^{-15}, 2^{-14}, \dots, 2^2, 2^3\}$, where C controls the strength of L2 regularization and γ is the RBF kernel coefficient. For MLP regression, we tuned over the number of hidden layers (one or two), the number of hidden units per layer (64 or 128), and the L2 regularization term $\alpha \in \{0.0001, 0.001, 0.01, 0.05, 0.1, 0.5, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.7, 1.9, 2.0, 3.0, 4.0, 5.0\}$.

To evaluate predictive performance during testing, we used three complementary performance metrics. First, we used RMSE as a measure of the distance between the predictions and labels in the same units as the labels. Given that the labels were standardized to have a standard deviation of 1, a perfect model would have RMSE = 0.0, a model always predicting the mean would have RMSE = 1.0, and a model with RMSE > 1.0 would be worse than a model always predicting the mean. Second, we used the coefficient of determination (R^2) as an estimate of the proportion of label variance explained by the features in a model. A perfect model would have $R^2 = 1.0$, a model always predicting the mean would have $R^2 = 0.0$, and a model with $R^2 < 0.0$ would be worse than a model always predicting the mean. Third, we used the Pearson correlation coefficient (r) as a standardized measure of the strength of the relationship between the predictions and labels. A perfect model would have $r = 1.0$, an uninformative model would have $r = 0.0$, and a model with $r < 0.0$ would be misleading. Unlike the other two metrics, correlation does not require that the predictions and labels be identical—only that they be linearly related.

Table 3: Performance of predictive models

Algorithm	Modality	RMSE	R^2	r
Elastic Net	Visual	0.89	-0.03	0.45
Elastic Net	Linguistic	0.76	0.30	0.65
Elastic Net	Multimodal	0.65	0.45	0.74
SVR	Visual	0.87	0.05	0.44
SVR	Linguistic	0.75	0.34	0.66
SVR	Multimodal	0.78	0.27	0.63
MLP	Visual	0.89	0.02	0.43
MLP	Linguistic	0.76	0.30	0.65
MLP	Multimodal	0.72	0.36	0.68

**Figure 4: Performance of Elastic Net models by modality**

To compare the performance of models trained using different features, we used non-parametric bootstrap resampling with percentile confidence intervals and p -values [18]. Specifically, we created 2000 resamples ($n = 160$) of the fold-level differences between models (through sampling with replacement) and, as in Table 3, calculated the median as a robust measure of central tendency.

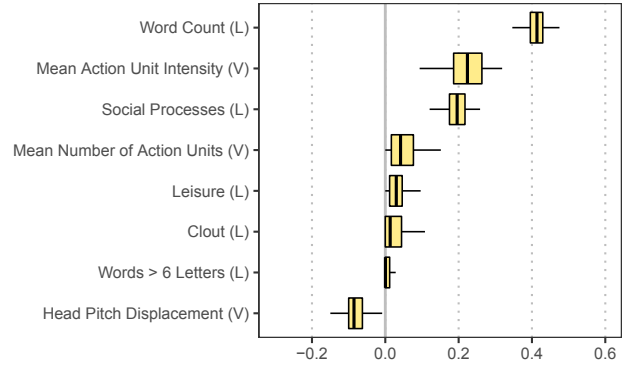
6.3 Predictive Modeling Results

Because we had eight test folds in each of the 20 repetitions of our cross-validation procedure, our experiment yielded a distribution of 160 scores on each performance metric. In Table 3, we present the median of each distribution as a robust measure of its central tendency. A full record of the results for all models is provided in the supplemental materials.

Research Question 1. To investigate our first research question about the prediction of emotional expressiveness from behavioral signals, we examined the best performing algorithm-modality combination: Elastic Net using multimodal behavioral signals as features. The predictions of this model were an average of 0.65 standard deviations from the corresponding label values, the features explained nearly half (45%) of the variance in the labels, and the predictions were highly correlated with the labels (0.74). The strength of the correlation in particular suggests that the predictions of this model could be used as a proxy for human ratings in settings where preserving

Table 4: Comparison of modalities for Elastic Net

Difference in RMSE	Δ	95% CI	p
Visual – Linguistic	0.16	[0.11, 0.18]	<.001
Visual – Multimodal	0.25	[0.22, 0.27]	<.001
Linguistic – Multimodal	0.10	[0.08, 0.11]	<.001
Difference in R^2	Δ	95% CI	p
Visual – Linguistic	-0.31	[-0.36, -0.19]	<.001
Visual – Multimodal	-0.46	[-0.53, -0.41]	<.001
Linguistic – Multimodal	-0.17	[-0.21, -0.15]	<.001
Difference in r	Δ	95% CI	p
Visual – Linguistic	-0.21	[-0.27, -0.14]	<.001
Visual – Multimodal	-0.30	[-0.34, -0.25]	<.001
Linguistic – Multimodal	-0.11	[-0.13, -0.09]	<.001

**Figure 5: Coefficients from multimodal Elastic Net models**

linear relationships is sufficient. However, the fact that more than half of the variance in the labels was unexplained by the features suggests that there is still more to emotional expressiveness than what our behavioral measures were able to capture.

We suspect that much of the unexplained variance is related to acoustic behavioral signals (e.g., speech prosody); the rest may be related to visual and linguistic signals that our representations did not capture (e.g., specific facial actions or gestures) or cultural and individual differences in patterns of expressive behavior.

Research Question 2. To investigate our second research question about the relative importance of the visual and linguistic modalities (given our selected behavioral representations), we compared the performance of the three Elastic Net models. The results of the modality comparison are provided in Table 4.

Our first finding was that the linguistic model significantly outperformed the visual model on all performance metrics. This result may suggest that emotional expressiveness is conveyed more through linguistic behavioral signals than through visual behavioral signals. Alternatively, it may suggest that our linguistic representations (based on manual transcripts) were more reliable than our visual representations (based on automatic tracking).

Our second finding was that the multimodal model significantly outperformed the the visual model and the linguistic model on all performance metrics. These results suggest that both modalities made unique contributions to the prediction and that both are worth including in future work on emotional expressiveness.

Research question 3. To investigate our third research question about the relationships between specific behavioral signals and emotional expressiveness, we examined the distributions of the regression coefficients across all tuned multimodal Elastic Net models. Figure 5 depicts the distributions with a nonzero median (and all distributions are depicted in the supplemental materials).

With a median coefficient of around 0.4, the total number of words spoken (*Word Count*) was the behavioral signal with the strongest relationship to emotional expressiveness. As language is a primary channel of human communication, it is not surprising that participants rated higher on emotional expressiveness tended to speak more. That is, speaking more likely gave participants additional opportunities to express themselves.

Two other behavioral signals had median coefficients around 0.2: the average magnitude of all facial movements (*Mean Action Unit Intensity*) and the proportion of words that fell in the *Social Processes* category (e.g., family and friends). In general, more intense facial movements correspond to stronger emotion [32], and since we define emotional expressiveness as the tendency to outwardly display emotion, it follows that participants who show stronger emotion are more expressive. This result is also consistent with our prior findings on the role of action unit intensity in momentary emotional expressiveness [38]. The relationship between the *Social Processes* linguistic signal and emotional expressiveness suggests a link between sociability and emotional expressiveness that is consistent with the correlation between emotional expressiveness and self-reported extraversion observed in this study and others [25, 36]. Specifically, individuals who socialize more (or who speak more about their social habits and activities) may tend to be more expressive not only of such activities but also of their emotions.

Finally, five other behavioral signals had median coefficients between 0.0 and 0.1 or between 0.0 and -0.1 . Participants rated higher on emotional expressiveness had more facial movements (*Mean Number of Action Units*), more words that fell in the *Leisure* category (e.g., cook, chat, movie), higher scores on the *Clout* dimension (i.e., speaking with confidence and authority), more words longer than six letters (*Words > 6 Letters*), and smaller vertical head motions (*Head Pitch Displacement*). Some of the relationships between these signals (i.e., number of action units and leisure words) are fairly similar in their interpretation to the relationships described in the previous paragraph. We also believe that the last finding—the negative coefficient for vertical head motion—may reflect that highly expressive participants spend more time talking and less time listening (as nodding is often a sign of listening [46]). The associations between emotional expressiveness and *Clout* and long words are more novel. These results point to possible positive relationships between emotional expressiveness and assertiveness (a facet of extraversion). Individuals who exhibit greater confidence in word choice, whether through more dominant or more advanced vocabulary, may tend to be more emotionally expressive.

7 DISCUSSION

In this paper, we added highly reliable transcripts of spoken words and human perceptual ratings of emotional expressiveness to an existing video database of unstructured social interactions. We proposed and tested a series of models encompassing traditional statistical modeling and machine learning to predict emotional expressiveness from visual and linguistic features. Encouragingly, we found that the multimodal statistical model (Elastic Net) performed reasonably well, explaining nearly half of the variance in the perceptual ratings. We then examined what this model learned in predicting emotional expressiveness from behavioral signals. Our results uncovered novel relationships between emotional expressiveness and behaviors related to traits like sociability, assertiveness, and confidence. Our most important visual features—those derived from action unit measures—converge with prior findings on the prediction and characterization of facial expressiveness in a very different context [38]. The consistency of these results suggests a persistent role of facial action units in signaling emotional expressiveness.

Our results should be considered in light of several points. First, the GFT dataset (like many others in affective computing) included only healthy young adults from a single geographic area, and it is unknown how well our findings would generalize to other populations. Second, the context of the participants’ behavior (i.e., the group formation task) was unstructured and more naturalistic than previous work, but the interactions still took place in a laboratory setting. Finally, the relatively small size of the dataset (in terms of both the number of participants and the amount of time each participant was observed) likely limited the performance our predictive models were able to achieve. These limitations can all be addressed through more research in additional populations and contexts, and we hope this paper will serve to encourage other interdisciplinary researchers to study emotional expressiveness.

In conclusion, we laid the foundation for the computational study of *emotional expressiveness*, which captures to the degree to which a person tends to outwardly display their emotions. We developed a method to collect human ratings of emotional expressiveness with high reliability and used latent variable modeling techniques to refine these ratings. We then constructed models to predict and characterize these ratings in terms of interpretable visual and linguistic behavioral signals. This line of research holds the potential to improve emotional inferences through personalized (i.e., expressiveness-aware) algorithms, develop more satisfying and efficient robots and virtual agents, and aid medical professionals in the assessment of psychiatric and neurological conditions.

ACKNOWLEDGEMENTS

This paper is based upon work supported by the National Science Foundation (Awards 1722822 and 1734868) and National Institutes of Health (Awards R01-AA015773, R01-MH096951, U01-MH116925, and U01-MH116923). Any opinions and recommendations expressed in this material are those of the authors alone.

SUPPLEMENTAL MATERIALS

All source code from this paper can be found at <https://osf.io/umzar/>. The Sayette GFT database [24] is freely available for academic purposes (to request access, view the wiki at <https://osf.io/7wcyz/>).

REFERENCES

- [1] American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders* (fifth ed.). American Psychiatric Association, Washington, DC.
- [2] Ruben C. Arslan, Matthias P. Walther, and Cyril S. Tata. 2020. Formr: A Study Framework Allowing for Automated Feedback Generation and Complex Longitudinal Experience-Sampling Studies Using R. *Behavior Research Methods* 52 (2020), 376–387.
- [3] Tadas Baltrušaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. OpenFace 2.0: Facial Behavior Analysis Toolkit. In *Proceedings of the 13th IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE, Xi'an, China, 59–66.
- [4] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez, and Seth D. Pollak. 2019. Emotional Expressions Reconsidered: Challenges to Inferring Emotion from Human Facial Movements. *Psychological Science in the Public Interest* 20, 1 (2019), 1–68.
- [5] Joan-Isaac Biel and Daniel Gatica-Perez. 2012. The youtube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs. *IEEE Transactions on Multimedia* 15, 1 (2012), 41–55.
- [6] Clément Chastagnol and Laurence Devillers. 2012. Personality Traits Detection using a Parallelized Modified SFFS Algorithm. In *Proceedings of the 13th Annual Conference of the International Speech Communication Association*. ISCA, Portland, OR, 266–269.
- [7] Ira Cohen, Nicu Sebe, Ashutosh Garg, Lawrence S. Chen, and Thomas S. Huang. 2003. Facial Expression Recognition from Video Sequences: Temporal and Static Modeling. *Computer Vision and Image Understanding* 91, 1–2 (2003), 160–187.
- [8] Jeffrey F. Cohn, Itir Onal Ertugrul, Wen-Sheng Chu, Jeffrey M. Girard, and Zakia Hammal. 2019. Affective Facial Computing: Generalizability across Domains. In *Multimodal Behavior Analysis in the Wild: Advances and Challenges*, Xavier Alameda-Pineda, Elisa Ricci, and Nicu Sebe (Eds.). Academic Press, Cambridge, MA, 407–441.
- [9] Paul T. Costa and Robert R. McCrae. 1992. *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Psychological Assessment Resources, Odessa, FL.
- [10] Alan Cowen, Disa Sauter, Jessica L. Tracy, and Dacher Keltner. 2019. Mapping the Passions: Toward a High-Dimensional Taxonomy of Emotional Experience and Expression. *Psychological Science in the Public Interest* 20, 1 (2019), 69–90.
- [11] Marcus Credé, Peter Harms, Sarah Niehorster, and Andrea Gaye-Valentine. 2012. An Evaluation of the Consequences of Using Short Measures of the Big Five Personality Traits. *Journal of Personality and Social Psychology* 102, 4 (2012), 874–888.
- [12] Jamie DeCoster, Anne-Marie R. Iselin, and Marcello Gallucci. 2009. A Conceptual and Empirical Examination of Justifications for Dichotomization. *Psychological Methods* 14, 4 (2009), 349–366.
- [13] Abhinav Dhall, Roland Goecke, Jyoti Joshi, Michael Wagner, and Tom Gedeon. 2013. Emotion Recognition in the Wild Challenge 2013. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*. ACM, Sydney, Australia, 509–516.
- [14] Abhinav Dhall, Amanjot Kaur, Roland Goecke, and Tom Gedeon. 2018. EmotiW 2018: Audio-Video, Student Engagement and Group-level Affect Prediction. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*. ACM, Boulder, CO, 653–656.
- [15] Abhinav Dhall, O. V. Ramana Murthy, Roland Goecke, Jyoti Joshi, and Tom Gedeon. 2015. Video and image based emotion recognition challenges in the wild: EmotiW 2015. In *Proceedings of the 17th ACM International Conference on Multimodal Interaction*. ACM, Seattle, WA, 423–426.
- [16] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. 2014. Decaf: A deep convolutional activation feature for generic visual recognition. In *Proceedings of the 31st International Conference on Machine Learning*. IMLS, Beijing, China, 647–655.
- [17] Harris Drucker, Christopher J. C. Burges, Linda Kaufman, Alex J. Smola, and Vladimir Vapnik. 1997. Support Vector Regression Machines. In *Advances in Neural Information Processing Systems*, Vol. 9. MIT Press, Cambridge, MA, 155–161.
- [18] Bradley Efron and Robert J. Tibshirani. 1993. *An Introduction to the Bootstrap*. Chapman and Hall, New York, NY.
- [19] Paul Ekman, Wallace V. Friesen, and John Hager. 2002. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Research Nexus, Salt Lake City, UT.
- [20] Jessica K. Flake, Jolynn Pek, and Eric Hehman. 2017. Construct Validation in Social and Personality Research: Current Practice and Recommendations. *Social Psychological and Personality Science* 8, 4 (2017), 370–378.
- [21] Howard S. Friedman, Louise M. Prince, and Ronald E. Riggio. 1980. Understanding and Assessing Nonverbal Expressiveness: The Affective Communication Test. *Journal of Personality and Social Psychology* 39, 2 (1980), 333–351.
- [22] Mauricio Garnier-Villarreal and Terrence D. Jorgensen. 2020. Adapting Fit Indices for Bayesian Structural Equation Modeling: Comparison to Maximum Likelihood. *Psychological Methods* 25, 1 (2020), 46–70.
- [23] Jeffrey M. Girard. 2020. agreement: An R Package for the Tidy Analysis of Agreement and Reliability. <https://github.com/jmgirard/agreement>
- [24] Jeffrey M. Girard, Wen-Sheng Chu, Laszlo A. Jeni, Jeffrey F. Cohn, F. De La Torre, and Michael A. Sayette. 2017. Sayette Group Formation Task (GFT) Spontaneous Facial Expression Database. In *Proceedings of the 12th IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE, Washington, DC, 581–588.
- [25] James J. Gross and Oliver P. John. 1995. Facets of Emotional Expressivity: Three Self-Report Factors and Their Correlates. *Personality and Individual Differences* 19, 4 (1995), 555–568.
- [26] Hatice Gunes and Björn W. Schuller. 2013. Categorical and Dimensional Affect Analysis in Continuous Input: Current Trends and Future Directions. *Image and Vision Computing* 31, 2 (2013), 120–136.
- [27] Kilem L. Gwet. 2014. *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement among Raters* (fourth ed.). Advanced Analytics, Gaithersburg, MD.
- [28] Adriana Hamacher, Nadia Bianchi-Berthouze, Anthony G. Pipe, and Kerstin Eder. 2016. Believing in BERT: Using Expressive Communication to Enhance Trust and Counteract Operational Error in Physical Human-Robot Interaction. In *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, New York, NY, 493–500.
- [29] Richard I. Hartley and Andrew Zisserman. 2004. *Multiple View Geometry in Computer Vision* (second ed.). Cambridge University Press, Cambridge, England.
- [30] Li-tze Hu and Peter M. Bentler. 1999. Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling* 6 (1999), 1–55.
- [31] Ashish Kapoor and Rosalind W. Picard. 2005. Multimodal Affect Recognition in Learning Environments. In *Proceedings of the 13th ACM International Conference on Multimedia*. ACM, Singapore, 677–682.
- [32] Dacher Keltner, Paul Ekman, Gian C. Gonzaga, and Jennifer Beer. 2003. Facial Expression of Emotion. In *Handbook of Affective Sciences*, H. Hill Goldsmith, Klaus R. Scherer, and Richard J. Davidson (Eds.). Oxford University Press, Oxford, England, 415–432.
- [33] Laura A. King and Robert A. Emmons. 1990. Conflict over Emotional Expression: Psychological and Physical Correlates. *Journal of Personality and Social Psychology* 58, 5 (1990), 864–877.
- [34] Rex Kline. 2015. *Principles and Practice of Structural Equation Modeling* (fourth ed.). Guilford Press, New York, NY.
- [35] Terry K. Koo and Mae Y. Li. 2016. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine* 15, 2 (2016), 155–163.
- [36] Ann M. Kring, David A. Smith, and John M. Neale. 1994. Individual Differences in Dispositional Expressiveness: Development and Validation of the Emotional Expressivity Scale. *Journal of Personality and Social Psychology* 66, 5 (1994), 934–949.
- [37] Shan Li and Weihong Deng. 2020. Deep Facial Expression Recognition: A Survey. *IEEE Transactions on Affective Computing* (2020).
- [38] Victoria Lin, Jeffrey M. Girard, and Louis-Philippe Morency. 2020. Context-Dependent Models for Predicting and Characterizing Facial Expressiveness. In *Proceedings of the AAAI-20 Workshop on Affective Content Analysis*, Vol. 2614. AAAI, New York, NY, 11–28.
- [39] Gwen C. Littlewort, Marian Stewart Bartlett, and Kang Lee. 2007. Faces of Pain: Automated Measurement of Spontaneous Facial Expressions of Genuine and Posed Pain. In *Proceedings of the 9th International Conference on Multimodal Interfaces*. ACM, Nagoya, Japan, 15–21.
- [40] Robert R. McCrae and Paul T. Costa. 1999. A Five-Factor Theory of Personality. In *Handbook of Personality: Theory and Research* (second ed.), L. A. Pervin and O. John (Eds.). Guilford Press, New York, NY, 139–153.
- [41] Kenneth O McGraw and S P Wong. 1996. Forming Inferences about Some Intraclass Correlation Coefficients. *Psychological Methods* 1, 1 (1996), 30–46.
- [42] Daniel McNeish and Melissa Gordon Wolf. 2020. Thinking Twice about Sum Scores. *Behavior Research Methods* (2020).
- [43] Edgar C. Merkle and Yves Rosseel. 2018. Blavaan: Bayesian Structural Equation Models via Parameter Expansion. *Journal of Statistical Software* 85, 1 (2018), 1–30.
- [44] Gelareh Mohammadi and Alessandro Vinciarelli. 2012. Automatic Personality Perception: Prediction of Trait Attribution based on Prosodic Features. *IEEE Transactions on Affective Computing* 3, 3 (2012), 273–284.
- [45] Claude Montacé and Marie-José Caraty. 2012. Pitch and Intonation Contribution to Speakers' Traits Classification. In *Proceedings of the 13th Annual Conference of the International Speech Communication Association*. ISCA, Portland, OR, 526–529.
- [46] Monisha Pasupathi, Laura L. Carstensen, Robert W. Levenson, and John M. Gottman. 1999. Responsive listening in long-married couples: A psycholinguistic perspective. *Journal of Nonverbal Behavior* 23, 2 (1999), 173–193.
- [47] James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, and Kate Blackburn. 2015. *The Development and Psychometric Properties of LIWC2015*. University of Texas at Austin, Austin, TX, USA.

- [48] Stavros Petridis and Maja Pantic. 2008. Audiovisual Discrimination between Laughter and Speech. In *Proceedings of the 33rd IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, Las Vegas, NV, 5117–5120.
- [49] Tim Polzehl, Sebastian Moller, and Florian Metze. 2010. Automatically Assessing Personality from Speech. In *Proceedings of the 4th IEEE International Conference on Semantic Computing*. IEEE, Pittsburgh, PA, 134–140.
- [50] Víctor Ponce-López, Baiyu Chen, Marc Oliu, Ciprian Corneanu, Albert Clapés, Isabelle Guyon, Xavier Baró, Hugo Jair Escalante, and Sergio Escalera. 2016. Chalearn lap 2016: First round challenge on first impressions-dataset and results. In *European Conference on Computer Vision*. Springer, Amsterdam, the Netherlands, 400–418.
- [51] Robert Rosenthal. 1982. Conducting Judgment Studies. In *Handbook of Methods in Nonverbal Behavior Research*, Klaus R. Scherer and Paul Ekman (Eds.). Cambridge University Press, New York, NY, 287–365.
- [52] Michael A. Sayette, Kasey G. Creswell, John D. Dimoff, Catharine E. Fairbairn, Jeffrey F. Cohn, Bryan W. Heckman, Thomas R. Kirchner, John M. Levine, and Richard L. Moreland. 2012. Alcohol and Group Formation: A Multimodal Investigation of the Effects of Alcohol on Emotion and Social Bonding. *Psychological Science* 23, 8 (2012), 869–878.
- [53] Björn Schuller, Ronald Müller, Benedikt Hörnler, Anja Höthker, Hitoshi Konosu, and Gerhard Rigoll. 2007. Audiovisual Recognition of Spontaneous Interest within Conversations. In *Proceedings of the 9th International Conference on Multimodal interfaces*. ACM, Nagoya, Japan, 30–37.
- [54] Björn Schuller, Stefan Steidl, Anton Batliner, Elmar Nöth, Alessandro Vinciarelli, Felix Burkhardt, Rob van Son, Felix Weninger, Florian Eyben, Tobias Bocklet, et al. 2012. The InterSpeech 2012 Speaker Trait Challenge. In *Thirteenth Annual Conference of the International Speech Communication Association*. ISCA, Portland, OR, 254–257.
- [55] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. 2014. CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, Columbus, OH, 806–813.
- [56] Jacopo Staiano, Bruno Lepri, Ramanathan Subramanian, Nicu Sebe, and Fabio Pianesi. 2011. Automatic modeling of personality states in small group interactions. In *Proceedings of the 19th ACM International Conference on Multimedia*. ACM, Scottsdale, AZ, 989–992.
- [57] Timothy J. Sullivan, Fea R. Leifker, and Amy D. Marshall. 2018. Observed Emotional Expressivity, Posttraumatic Stress Disorder Symptoms, and Intimate Partner Violence Perpetration among Community Couples. *Journal of Traumatic Stress* 31, 3 (2018), 352–361.
- [58] Auke Tellegen, David T. Lykken, Thomas J. Bouchard, Kimerly J. Wilcox, Nancy L. Segal, and Stephen Rich. 1988. Personality Similarity in Twins Reared Apart and Together. *Journal of Personality and Social Psychology* 54, 6 (1988), 1031–1039.
- [59] Yan Tong, Wenhui Liao, and Qiang Ji. 2007. Facial Action Unit Recognition by Exploiting their Dynamic and Semantic Relationships. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29, 10 (2007), 1683–1699.
- [60] Panagiotis Tzirakis, George Trigeorgis, Mihalis A. Nicolaou, Björn W. Schuller, and Stefanos Zafeiriou. 2017. End-to-End Multimodal Emotion Recognition using Deep Neural Networks. *IEEE Journal of Selected Topics in Signal Processing* 11, 8 (2017), 1301–1309.
- [61] Fabio Valente, Samuel Kim, and Petr Motlicek. 2012. Annotation and Recognition of Personality Traits in Spoken Conversations from the AMI Meetings Corpus. In *Proceedings of the 13th Annual Conference of the International Speech Communication Association*. ISCA, Portland, OR, 1183–1186.
- [62] Michel F. Valstar, Björn W. Schuller, Jarek Krajewski, Roddy Cowie, and Maja Pantic. 2014. AVEC 2014: The 4th International Audio/Visual Emotion Challenge and Workshop. In *Proceedings of the ACM International Conference on Multimedia*. ACM, Orlando, FL, 1243–1244.
- [63] Laurence Vidrascu and Laurence Devillers. 2005. Detection of Real-Life Emotions in Call Centers. In *Proceedings of the 9th European Conference on Speech Communication and Technology*. ISCA, Lisbon, Portugal, 1841–1844.
- [64] Alessandro Vinciarelli and Gelareh Mohammadi. 2014. A Survey of Personality Computing. *IEEE Transactions on Affective Computing* 5, 3 (2014), 273–291.
- [65] Aidan G. C. Wright. 2014. Current Directions in Personality Science and the Potential for Advances through Computing. *IEEE Transactions on Affective Computing* 5, 3 (2014), 292–296.
- [66] Zhihong Zeng, Yuxiao Hu, Glenn I. Roisman, Zhen Wen, Yun Fu, and Thomas S. Huang. 2007. Audio-visual Spontaneous Emotion Recognition. In *Artificial Intelligence for Human Computing*. Springer, New York, NY, 72–90.
- [67] Zhihong Zeng, Maja Pantic, Glenn I. Roisman, and Thomas S. Huang. 2009. A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 1 (2009), 39–58.
- [68] Hui Zou and Trevor Hastie. 2005. Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67, 2 (2005), 301–320.