

# An Online Attachment Style Recognition System Based on Voice and Machine Learning

Lucía Gómez-Zaragozá , Javier Marín-Morales , Elena Parra Vargas , Irene Alice Chicchi Giglioli , and Mariano Alcañiz Raya

**Abstract**—Attachment styles are known to have significant associations with mental and physical health. Specifically, insecure attachment leads individuals to higher risk of suffering from mental disorders and chronic diseases. The aim of this study is to develop an attachment recognition model that can distinguish between secure and insecure attachment styles from voice recordings, exploring the importance of acoustic features while also evaluating gender differences. A total of 199 participants recorded their responses to four open questions intended to trigger their attachment system using a web-based interrogation system. The recordings were processed to obtain the standard acoustic feature set eGeMAPS, and recursive feature elimination was applied to select the relevant features. Different supervised machine learning models were trained to recognize attachment styles using both gender-dependent and gender-independent approaches. The gender-independent model achieved a test accuracy of 58.88%, whereas the gender-dependent models obtained 63.88% and 83.63% test accuracy for women and men respectively, indicating a strong influence of gender on attachment style recognition and the need to consider them separately in further studies. These results also demonstrate the potential of acoustic properties for remote assessment of attachment style, enabling fast and objective identification of this health risk factor, and thus supporting the implementation of large-scale mobile screening systems.

**Index Terms**—Acoustic features, artificial intelligence, attachment, gender, psychometrics, speech analysis, statistical machine learning, voice.

## I. INTRODUCTION

ATTACHMENT theory is a wide-ranging social development theory introduced by John Bowlby that describes

Manuscript received 6 April 2023; revised 5 July 2023; accepted 4 August 2023. Date of publication 11 August 2023; date of current version 7 November 2023. This work was supported in part by the Generalitat Valenciana under Grant ACIF/2021/187, and its funded Project Mixed reality and brain decision - REBRAND under Grant PROMETEO/2019/105, and in part by the Universitat Politècnica de València under Grants PAID-10-20 and PAID-PD-22. (Corresponding author: Lucía Gómez-Zaragozá.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Research Ethics Committee of the Polytechnic University of Valencia Application No. P01\_08\_07\_20.

The authors are with the Instituto Universitario de Investigación en Tecnología Centrada en el Ser Humano, Universitat Politècnica de València, 46022 Valencia, Spain (e-mail: lugoza@htech.upv.es; jamarmo@htech.upv.es; elparvar@htech.upv.es; alice.chicchi@gmail.com; malcaniz@htech.upv.es).

Digital Object Identifier 10.1109/JBHI.2023.3304369

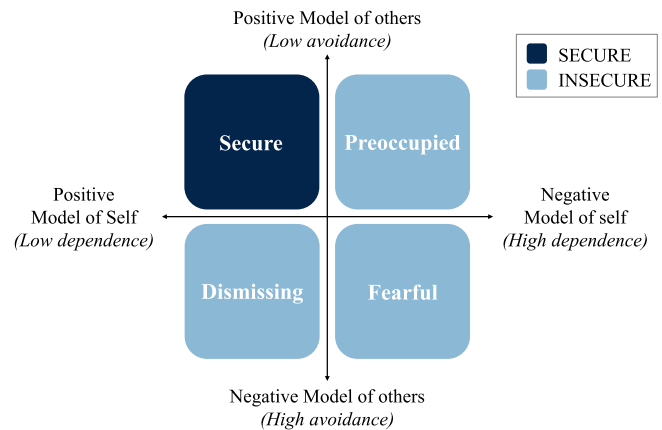


Fig. 1. Four-category model of adult attachment. Adapted from [5].

the origin of the patterns that take place in close interpersonal relationships, known as attachment styles [1]. Bowlby proposed the attachment behavioral system as a psychological organization that regulates behaviors that are necessary for acquiring and maintaining stable and valuable emotional relationships across the lifespan. The theory states that throughout their early stages of life, children develop attachment behaviors in the form of basic emotional expression as a mechanism to obtain the closeness of their attachment figures (typically their parents) during uncertain and stressful situations [1]. Depending on how the attachment figures respond, children adapt their behaviors and construct what Bowlby called internal working models of the self and others [2]. The self-model refers to the mental representation of oneself and one's worth, whereas the other-model consists of the expected response of attachment figures to one's behavior. According to Bowlby [2], the internal working models formed in childhood tend to be stable in adulthood due to the predisposition to face new situations from preexisting models. This stability can be reinforced by other factors such as a secure environment, resulting in a stable way of relating to the world in adulthood [2], [3]. Attachment style in adulthood was categorized by Bartholomew [4] into regions in a two-dimensional space consisting of the model of self and the model of others, which can also be conceptualized in terms of dependency and avoidance of the relationship respectively [5]. As illustrated in Fig. 1, their combination results in four types of attachment style. Secure attachment corresponds to those who

see themselves as deserving of love and support, and others as trustworthy and accessible. In insecure-fearful attachment, individuals have a negative self-image and see others as unreliable and rejective. Insecure-preoccupied individuals have a negative view of themselves but a favorable impression of others. Finally, insecure-dismissing category includes subjects characterized by a sense of worthiness but a negative representation of others, avoiding close relationships to protect themselves from being disappointed.

Attachment styles have been significantly associated with mental and physical health. Secure attachment is the basis for achieving good mental health [6], whereas insecure attachment leads individuals to higher risk of suffering from mental disorders [7]. Specifically, the latter has been associated with psychopathologies such as depression, generalized anxiety disorder and borderline personality, as well as some symptoms like stress, anxiety and eating disorders [7]. Concerning physical health, insecure attachment styles have been linked with certain physiological states with a negative impact on the individual's physical condition, including elevated vagal nerve tone and higher cortisol levels [8]. In fact, a study investigating the association between attachment patterns and health outcomes found that insecure attachment may be a risk factor for developing chronic diseases [9]. Their findings support that individuals with avoidant attachment are positively associated with pain-related conditions, such as frequent or severe headaches. On the other hand, individuals with anxious attachment were more likely to experience a broader range of health conditions, particularly those involving the cardiovascular system, such as high blood pressure, heart attack, and stroke. Notably, the study showed no association between secure attachment and any of the examined health conditions. Another factor influencing physical and mental health is lifestyle: adults with secure attachment style adopt healthier preventative health behaviors, including healthy diet and exercise, than those with insecure attachment style [10].

Recent research has suggested that gender differences need to be considered when exploring attachment theory [11]. Attachment behaviors during childhood are explained mainly in neutral terms since infants from both sexes face the same purpose of obtaining closeness and protection from attachment figures. Gender differences begin to be observed during middle childhood (e.g. [12], [13], [14]); but it is not clear if they emerge for the first time at that age or if they are revealed because they become more pronounced. These differences continue in the adult stage: it has been reported that whereas women tend to be higher in anxiety, men tend to be higher in avoidance [15], although the magnitude of the dissimilarity varies from small to moderate depending on the cultural region considered [16]. Other investigation [17] found significant effects in romantic attachment when anxiety and avoidance dimensions were divided into facets: avoidance into self-reliance and discomfort with closeness, and anxiety into neediness, preoccupation, and a rejected desire for closeness. Self-reliance and rejected desire for closeness were higher in men whereas preoccupation and neediness were higher in women; discomfort with closeness was similar for both genders. Moreover, attachment styles influence parenting and mating in adulthood, so from an evolutionary

perspective, it is reasonable to consider differences between the two sexes [11].

The assessment of adult attachment styles is mainly done through self-report questionnaires and interviews, as detailed in [18]. Self-report measures are reliable and inexpensive but can be criticized for being too abstract, decontextualized and affected by social desirability bias. Interviews, on the other hand, achieve less response bias and can facilitate the activation of the attachment system but require more time (both for administration and coding), specialized training in coding and may be influenced by the examiner's interpretation. Moreover, Social Cognitive Neuroscience research has rejected social cognition models suggesting that humans can accurately analyze and verbalize their emotions, attitudes, and behaviors [19]. Instead, studies indicate that social interactions are largely regulated by unconscious processes [20]. Traditional attachment psychometrics evaluate individuals' feelings and behaviors directly from their conscious verbal responses, as only the transcripts are used for coding the most commonly used interview, the Adult Attachment Interview (AAI) [21]. Computational Psychiatry is a growing field that aims to provide accurate quantitative models between psychophysiological indicators (called implicit biomarkers) and explicit behaviors and responses. Implicit measures can capture unconscious brain processes, thus constituting a powerful alternative for detecting the attachment manifestation in the individual's unconscious response. Furthermore, since attachment theory is based on the quality of interpersonal relationships, a relevant biomarker in attachment is represented by the voice. As a means of human communication, the voice is one of the most natural and versatile, capable of expressing subjective ideas and conveying both linguistic and emotional information [22]. Extensive research has evidenced a link between variations in acoustic parameters and different emotions, establishing a growing research area called Speech Emotion Recognition [23]. Psychological conflicts and traumatic episodes can also affect the voice [24]. In some cases, traumatic experiences can even lead to clinical voice disorders like psychogenic aphonia, which prevents the person from speaking without any apparent physical cause. Therefore, it could be inferred that the voice's physiology might be systematically influenced by psychological factors, and particularly by the experiences occurring during early emotional development. Consequently, a relationship might be established between attachment and voice: since attachment-style (especially insecure attachment) affects physiological factors such as skin conductance, heart rate, or brain activity, it may also affect the voice.

Previous studies investigating attachment style and acoustic features [25], [26], [27], [28], [29], [30], [31], [32], detailed in the following section, have several limitations. Primarily, most of them included samples that were mostly or exclusively women [26], [27], [28], [29], [30], with small sample sizes [25], [26], [27], [28], [29], [30]. A recent article [32] studied a slightly larger sample, but did not provide its gender distribution. Only one study included a larger sample balanced for gender [31], and found differences between women and men, indicating the need to explore gender differences in attachment style. Regarding speech sample collection, several studies used limited

responses that might not accurately reflect typical speech [25], [26], [29], [31], and they generally did not include tasks related to attachment [25], [26], [31]. Therefore, it is possible that the participants' attachment system was not triggered, potentially compromising the results. Furthermore, all previous investigations [25], [26], [27], [28], [29], [30], [31], [32] were carried out in laboratory settings, which could also have influenced the outcomes. Finally, most of the previous investigations [25], [26], [27], [29], [30], [31] used statistical analysis to explore the relationship between attachment style and acoustic features, with only two studies [28], [32] exploring machine learning models, indicating the need for further research to validate the generation of attachment style recognition models using speech.

The aim of this research is to create an online attachment recognition model to distinguish between secure and insecure attachment styles using acoustic features extracted from voice recordings. To this end, three specific objectives are defined: 1) the collection of a large and balanced sample in terms of gender and attachment style; 2) the use of open-ended questions intended to trigger the attachment system while providing natural-speaking samples; 3) the implementation of online sample collection using mobile devices or computers, for future implementations of the system remotely. In addition, gender differences in attachment manifestation through voice are investigated by creating gender-independent and gender-dependent models for attachment recognition. For this purpose, we gathered a sample of 199 participants through an online system, where they recorded their answers to four open-ended questions. Subsequently, we extracted a set of standard acoustic features from each audio. Lastly, we employed feature selection techniques and trained various machine learning models. To the best of our knowledge, no previous work has developed a machine learning model to recognize attachment style from typical speech using a remote system with attachment-related questions. Therefore, the main novelty of the work is that it represents a pioneering effort in collecting speech recordings for online attachment style recognition. This makes it the first study to propose a system that can be deployed remotely outside of a laboratory setting. Moreover, we provide insights regarding the role of gender and the impact of acoustic features in the recognition of attachment styles.

This article is structured as follows. Section II describes the previous work related to attachment style and speech. Section III presents the database collected. Section IV includes the methodology applied to extract the acoustic features and train the models. Section V presents the results. Section VI discusses the results, compares them with previous work and describes the limitations and implications. Finally, Section VII concludes the research and outlines the main findings.

## II. PREVIOUS WORK

Previous studies have investigated the relationship between individuals' vocal cues and their attachment style. In [25], prosodic parameters were extracted from the voice range profile of a group of singers, and statistically significant differences were found between some of these features and anxious and

TABLE I  
COMPARISON OF THE APPROACHES USED IN THE LITERATURE TO GATHER SPEECH SAMPLES TO EVALUATE ATTACHMENT

Ref.	Sample	Methodology	Context	Setting	ML
[25]	$N_1 = 25, N_2 = 26$ (W:14, M:11) <sub>1</sub> (W:20, M:5, O:1) <sub>2</sub> N = 29	Voice range profile	No	Lab	No
[26]	(W:29, M:0)	Reading, sustained vowel, and picture description	No	Lab	No
[27]	N = 59 (W:45, M:14)	Biometric Attachment Test (BAT)	Yes	Lab	No
[28]	N = 76 (W:55, M:21)	Biometric Attachment Test (BAT) v2	Yes	Lab	Yes
[29]	N = 67 (W:67, M:0)	Question 3 of the AAI	Yes	Lab	No
[30]	N = 10 (W:10, M:0)	First half of the AAI	Yes	Lab	No
[31]	N = 206 (W:124, M:82)	Speaking 5 words	No	Lab	No
[32]	N = 118 (W: n/a, M: n/a)	Problem solving discussion	Yes	Lab	Yes

avoidant attachment, respectively. Becker [26] investigated the relationship between the variation in acoustic features extracted from three different tasks and the scores derived from a self-report questionnaire about traumatic experiences; however, no significant correlations were found. Parra et al. [27] applied the Biometric Attachment Test, an instrument for automatic attachment evaluation based on different psychophysiological measures (including voice), and composed of stimuli designed to trigger the attachment system. Later, in [28], the multimodal features automatically extracted with Biometric Attachment Test were used to create a machine learning (ML) model to predict the attachment style. Paralinguistic features extracted from voice recordings were shown to be fundamental in the detection of avoidance attachment. Other works [29], [30] explored acoustic parameters extracted from audio segments corresponding to words recorded in an attachment assessment context. Spinelli et al. [29] analyzed the words used by the individuals to describe the relationship with their mother/father during childhood – a question included in the AAI – whereas Moneta et al. [30] focused on the word “Mutter” (mother) when it was pronounced during the first half of the AAI. Nevertheless, these investigations related the differences in acoustic parameters to emotional activation rather than emphasizing their direct association with attachment style. More recently, Zhang and Zheng [31] analyzed vocal properties from individuals pronouncing five words and found correlations of men's parameters with anxious attachment, whereas no correlations were found for women. Koçak et al. [32] investigated the prediction of secure vs. fearful attachment style during two problem solving discussions of recently married couples. They used low-level acoustic features, i-vectors and sentiment analysis from manual transcriptions to train machine learning models, and obtained higher performances using only low-level acoustic features.

Some limitations of the previous investigations are summarized in Table I. Almost all of the studies [26], [27], [28], [29], [30] presented samples that were mostly or entirely represented by women, ranging from 10 to 76 subjects. The first sample used in [25] was gender-balanced but consisted only of singers, so more research is needed to analyze whether the results can be extrapolated to the general population. Exceptionally the sample

in [31] consisted of 206 participants, including 124 women, and they found differences in the results for both genders. The sample in [32] initially comprised 206 individuals, but models were finally trained with 118 participants, and no information was provided on the gender distribution of this final set. They found that the performance of the models was higher when they included features of both genders to predict the attachment style compared to each gender independently, with no significant differences. It is therefore necessary to assess whether there are indeed gender differences in vocal cues based on attachment style, as also suggested in previous research [11], [15], [17].

Concerning speech sample collection, some investigations [25], [26], [29], [31] used tasks like reading, sustaining a vowel or pronouncing individual words. These procedures facilitate sample comparison between subjects, but such a limited response may not be representative of usual speech. More open answers can be obtained from the verbal description of a picture, as in [26], although this task does not resemble conversational speech. Similarly, in studies by Parra et al., subjects were asked to verbally describe their feelings after stimuli exposure [27], [28], but the laboratory environment and the awareness that there was an expected answer could have influenced their speech. A more natural-speaking approach was presented by Moneta et al., who used sample recordings of interview responses [30]. However, the analysis was performed at the word level, as the authors restricted the recordings to segments where the subject pronounced a specific word. Koçak et al. investigated dialogues between recently married couples during two problem solving discussions [32]. They used the recordings from the full sessions, but they required manual processing to divide each dialogue in turns, remove segments with overlapping voices and exclude some of the turns to balance the number of samples in training. Consequently, further investigation is required to examine tasks that encourage reflective and open answers, enabling more comprehensive analysis and ultimately enhancing their validity.

Regarding methodology, it is important to consider that attachment behavior is a state-dependent trait, in that it appears in a consistent pattern only when the attachment system is activated. This activation occurs in specific situations where the individual is exposed to physical or psychological threats, including separation from attachment figures or isolation from others [1], [2]. Consequently, the studies found in the literature can be categorized according to their selection of procedures designed to trigger the attachment system. Some investigations [25], [26], [31] used tasks such as reading, sustaining a vowel, pronouncing individual words or describing a general picture aloud. Those methodologies do not expose the subject to psychological threats or stress contexts; therefore individuals could have difficulties in activating the attachment system [33]. Conversely, other studies [27], [28], [29], [30], [32] used speech samples from attachment assessment contexts, thereby theoretically achieving attachment system activation. It should be noted that attachment system activation can be facilitated when the individual is alone [1]. This factor was only considered in [27], [28], where the Biometric Attachment Test was administered to participants left alone in the room. In [32], the married couples were also left alone in a room for the discussion. However, all the studies

were conducted in laboratory settings, where the subject is often influenced by the non-familiar environment and the presence of researchers, which can imply a bias in the responses [34]. In contrast, through applications on mobile devices that can be used while alone at home, researchers may be able to overcome the mentioned limitations.

Finally, in most previous investigations the relationship between attachment style and acoustic features has been studied from a statistical view [25], [26], [27], [29], [30], [31]. Nevertheless, the small and unbalanced number of samples has limited their statistical power in some cases. To our knowledge, only two prior studies [28], [32] have explored the use of machine learning models. In [28], they developed an automatic scoring algorithm to measure attachment from multimodal features (including voice cues) collected during the Biometric Attachment Test. Paralinguistic features extracted from speech were critical for the detection of avoidance attachment, with a performance loss of 70% when they were removed from the model, whereas for the anxiety detection, the performance decreased by 12%. These results suggest a difference between unconscious speech behaviors for both groups, which raises, in turn, the possibility of using only acoustic parameters to predict attachment style. In addition, due to the small and gender-unbalanced data used for both the development of the Biometric Attachment Test and the scoring algorithm, and the lab-based conditions used, further research is needed to validate the generation of automatically assessed attachment style models using speech, and, in particular, to analyze the role of gender in this framework. In [32], they used statistics of low-level acoustic features, *i*-vectors and sentiment scores to train machine learning models (including decision trees, SVM and convolutional networks) to predict secure vs. fearful attachment style in couple interactions. They found that sentiment scores did not improve the models' accuracy, likely due to errors when translating the text, and neither did *i*-vectors, since low-level acoustic features contain the same information. Moreover, they obtained better results by including the spouse's features when predicting their partner's attachment style, compared to considering only their own features, claiming that spouses can influence each other during the conversation. However, the differences found were not significant, so further studies are needed to understand the role of gender in attachment prediction.

### III. DATASET

This section is organized into three parts. Firstly, it provides information regarding the study participants. Secondly, it presents the measures collected during the experimental procedure. Lastly, it describes the method applied to obtain the labels used in the models.

#### A. Participants

199 subjects participated in the study (gender: 51% women, 49% men; age: 35% in range 18–28, 25% in range 29–39, 23% in range 40–49, 18% in range 50–59). Participants were recruited according to specific inclusion criteria: being aged

between 18 and 59 and being Spanish speakers. The experimental protocol was approved by the Research Ethics Committee of the Polytechnic University of Valencia (reference number P01\_08\_07\_20).

## B. Measures

1) *Relationship Questionnaire*: Attachment style was assessed using the Spanish version of the Relationship Questionnaire, an instrument based on the four-category model described by Bartholomew and Horowitz [5] comprising two parts. First, the subject chose one out of four paragraphs that most closely matched the way they he/she was. Paragraphs consisted of general descriptions related to interpersonal relationships, each corresponding to a prototypical behavior of the secure, dismissing, preoccupied, and fearful attachment style. Second, the subject used a 7-point Likert scale to reflect the degree of conformity with each description.

2) *Audio Data*: The study's sample collection was carried out through an online access panel, which allowed participants to complete the study via their mobile phone or computer at the time they preferred. They were instructed to complete the study while alone in a room, which facilitates the activation of their attachment system [1]. To initiate the study, participants were required to accept the informed consent through the web-based system. Subsequently, the app presented the first part of the Relationship Questionnaire to the participants. Then, they recorded their responses to four open-ended questions: two of them focused on internal objectives (Q1 and Q2), one question about the social context (Q3) and a last question presenting a hypothetical situation to the participant (Q4). They were intended to trigger the attachment system following the guidelines presented below:

- Q1: *What are your goals in life? What do you have in mind to achieve your goals?* This question assesses the individual's reflective capacity and mental rigidity, two aspects also evaluated in the gold standard AAI [21].
- Q2: *What problems do you find in achieving your goals and how do you think you could solve them?* It is related to question six of the AAI [21], which states: "When you were upset as a child, what did you do, and what would happen?". Both questions aim to make the individual think about how to cope with adverse situations.
- Q3: *How can you define your relationship with others? If you could change a few things about the relationships you have, how would you like it to be?* It is closely related to attachment theory [1], [2], [3] itself, as it asks directly about the individual's interpersonal relationships.
- Q4: *Imagine that you were in a borderline situation where your child/lover was emotionally attacked (humiliated in a way that he or she was not aware of), what would you do?* Based on question 17 and 18 of the AAI [21], this question poses a borderline situation for the participant to reflect on his behaviour as a caregiver for other's protection, reminding him related childhood memories.

Finally, they completed the second part of the Relationship Questionnaire by clicking on the corresponding score under each paragraph. We decided to administer this stage after the

open-ended questions because the scores questions were designed to trigger the individual's attachment system, which could remain activated when answering the second part of the questionnaire and therefore overcome the limitation of self-report questionnaires. Noise conditions were checked manually by an expert listening to each recording. Participants were required to complete the survey in quiet places, so those subjects whose audio samples were considered noisy were eliminated from the study.

## C. Data Labeling

Participants were categorized into attachment categories based on the highest rating among the four attachment prototypes selected in the Relationship Questionnaire. If two or more prototypes received an equal rating, the option selected in the first part of the questionnaire was used to make a choice. However, if there was a three-way tie for the highest rating, the participant was excluded from the study. Then, the insecure attachment styles (preoccupied, dismissing and fearful) were grouped together as insecure category to study the binary classification against secure attachment. Additionally, self-model and other-model dimensions were obtained from the scores to each paragraph of the RQ for exploratory purposes only. Self-model classification was obtained by formulating (secure + dismissing) - (fearful + preoccupied). The other-model was calculated by computing (secure + preoccupied) - (fearful + dismissing). The scores for both dimensions were normalised to the range [0, 1].

## IV. METHODS

This section comprises two parts. First, it details the method for extracting acoustic features from collected data. Then, it describes the approach for feature selection and machine learning. An overview of the pipeline is provided in Fig. 2.

### A. Acoustic Feature Extraction

The speech recordings were collected in 16-bit mono WAV format, with sample rates of 44.1 kHz and 48 kHz depending on the user's recording device. First, a processing step was applied: audio signals were resampled to 44.1 kHz and then normalized to the range from  $-1$  to  $1$ . Then, with the aim of using a standard feature set that facilitates the understanding and reproducibility of the results, the extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS) [35] was extracted for each audio file using the open-source Speech and Music Interpretation by Large-space Extraction (openSMILE) toolkit [36]. Features in the eGeMAPS set were originally selected based on their sensitivity to capture changes in the voice produced by affective processes. It includes functionals (mainly arithmetic mean and coefficient of variation) applied to different low-level descriptors of the frequency, energy and spectral domain considering voiced and/or unvoiced regions, plus temporal features and the equivalent sound level. Additionally, the audio duration was included here, as the length of the participant's response may be relevant to this particular study, resulting in 89 features per audio. A

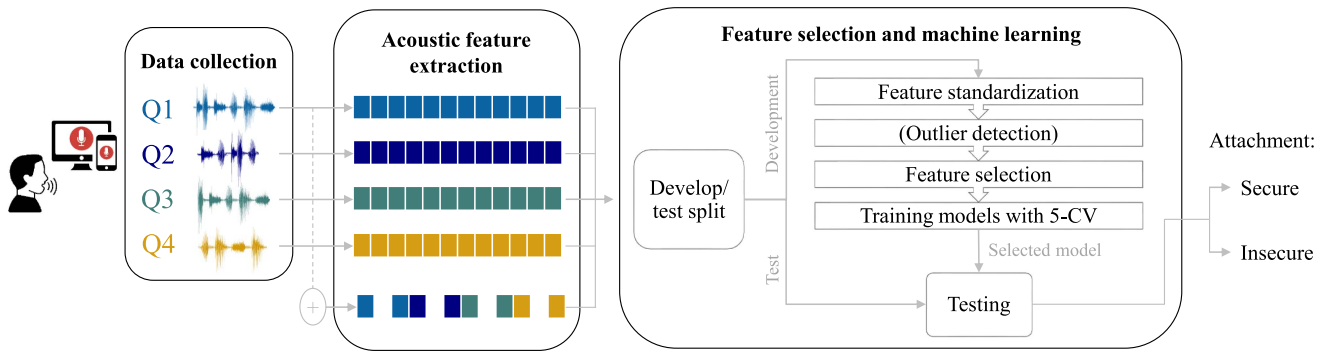


Fig. 2. Overview of the steps followed to create the attachment style recognition system.

TABLE II  
ACOUSTIC FEATURES CALCULATED FOR EACH RECORDING

Category	Features
Frequency	Jitter
	Fundamental frequency (F0)
	Formant 1, 2, 3 (frequency, bandwidth)
Energy / amplitude	Shimmer
	Loudness
	Harmonic-to-Noise Ratio (HNR)
Spectral	Alpha Ratio
	Hammarberg Index (HI)
	Spectral slope (0-5kHz, 0.5-1.5kHz)
	Formant 1, 2, 3 (relative energy)
	Harmonic difference H1-H2
	Harmonic difference H1-A3
	1-4 Mel-Frequency Cepstral Coefficients (MFCC)
Spectral flux	
Temporal	Rate of Loudness peaks
	Length of voiced regions
	Length of unvoiced regions
	Pseudo Syllable Rate
Other	Equivalent sound level
	Duration

summary of the acoustic features is shown in Table II and details on their calculations are described in [35].

### B. Feature Selection and Machine Learning

First, the dataset was partitioned into two sets: 157 participants for development ( $\approx 83\%$ ) and the remaining 32 for testing ( $\approx 17\%$ ). The test set included the same number of participants of each gender and attachment style, therefore preserving the ratios presented in the entire sample. Features were standardized by subtracting the mean and dividing by the standard deviation; both statistics were computed for the training set and independently for each feature. Then, the Isolation Forest [37] method was used to detect outliers within the acoustic features. It is a popular unsupervised detection method which uses a forest of random trees to isolate observations and measure their normality, where samples with shorter path lengths are highly likely to be anomalies. The method uses a contamination parameter that represents the expected proportion of outliers in the data, which

was set to 0.05 here in order to maintain the 95% of the distribution. The inclusion or not of the outliers removal was explored, including both options in the pipeline. Next Recursive Feature Elimination (RFE) [38] was applied for feature selection. RFE uses a cross-validation (CV) strategy to compute a criterion function for a given machine learning algorithm. Iteratively, one feature is eliminated at a time to create  $n-1$  subsets, for each of which the criterion function is calculated. The feature that helped the least in the classification is eliminated, that is, the feature that maximizes the criterion function when it is removed from the feature set. This process is repeated until a selected number of features are left. In this work, two features were removed at each iteration to reduce the computational time and a minimum of 10 features was established. Moreover, we used accuracy as the criterion function for a linear SVM and two options for the cross-validation strategy: the stratified K fold cross-validation with five folds, for the models trained for each open-ended question independently; and group k fold with five folds, when the concatenation of all of them was used to ensure gender-independent partitions. With the selected features, different machine learning algorithms were trained for classification between secure and insecure attachment style: Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF) and K-Nearest Neighbors (KNN). In this step, the same cross-validation strategy as in RFE was used to select the best model hyperparameters among the different combinations shown in Table III, following standard grid-search procedures [39]. Finally, testing was implemented using a bootstrap approach, with 50 repetitions using different random states for sampling the test with replacement.

To evaluate the models, the accuracy was computed as the proportion of correctly classified predictions compared to the total number of instances. As described in [40], accuracy is the most commonly used evaluation metric for classification problems, but it has limited discriminative power, especially with unbalanced data. The database used here is balanced in terms of secure and insecure attachment style. However, for a more comprehensive and detailed evaluation of the machine learning model's performance, the area under the ROC curve (AUC), sensitivity and specificity were also calculated as described in [40]. Insecure attachment was considered the relevant/positive element for sensitivity and specificity calculations. All steps were implemented using the Scikit-learn Python library.

TABLE III  
HYPERPARAMETER SET FOR EACH MACHINE LEARNING MODEL

Model	Hyperparameters
SVM	Kernel: Linear, Radial Basis Function (RBF), Sigmoid Gamma: 0.001, 0.01, 0.1, 1, $1/N$ , $1/(N \times V)$ C: 1, 10, 100, 1000
LR	Penalty: L2 Maximum iterations: 200 C: 20 logarithmically scaled values between -4 and 4 Solver: Library for Large Linear Classification (LIBLINEAR), Limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS)
RF	Bootstrapping samples: True, False Maximum depth of the tree: 10, 20, 30, None Number of trees in the forest: 100, 200, 400
KNN	Number of neighbors: 1, 3, 5, 7 Weight function: uniform, inverse of distance Metric for distance: Euclidean, Manhattan, Minkowski

$N$  = number of features,  $V$  = variance of the dataset.

The pipeline was replicated three times to explore different data combinations considering gender. First, a single model was created using the complete dataset. Second, only women participants were selected from the dataset, and the procedure was applied to obtain a new machine learning model. Finally, a third model was created for men subjects. Moreover, for each data combination, the features extracted for each open-ended question were analyzed independently, creating a classification model using each question, as well as for the concatenation of the features of all of them.

Additionally, to assess whether the performance of the trained models was above chance level, a permutation test was conducted. This involved randomly shuffling the assignments of samples to target values to create a scenario where there was no association input and output. The machine learning pipeline described before was then executed with these new targets to obtain an accuracy value for the test partition. This process was repeated 50 times to obtain a more accurate estimate. Finally, a one-sided t-test was used to analyse if the mean accuracy of the test bootstrapping is higher than the mean accuracy of the permutation test. The level of significance was set at  $\alpha = 0.05$ .

## V. RESULTS

This section is divided in two parts. First, the results from the Relationship Questionnaire and the final data distribution are presented. Then, the classification results are reported.

### A. Labeling Distribution

As described in Section III-C, participant's labels were obtained using the Relationship Questionnaire. Ten subjects were eliminated from the study due to a tie between ratings of different attachment styles, resulting in 189 participants with the distribution of gender, age and attachment style presented in Table IV. Fig. 3 shows the distribution of the attachment styles and the self-model and other-model dimensions.

### B. Classification Results

Fig. 4 shows the cross-validation results for the three different database configurations used to create the machine learning models: complete, women and men. Results have been separated

TABLE IV  
SOCIODEMOGRAPHIC INFORMATION OF THE 189 PARTICIPANTS

Attachment	Gender	Age range				Total
		18-28	29-39	40-49	50-59	
Secure	Women	16	13	11	11	51
	Men	13	14	15	11	53
Insecure	Women	23	10	7	5	45
	Men	14	12	8	6	40

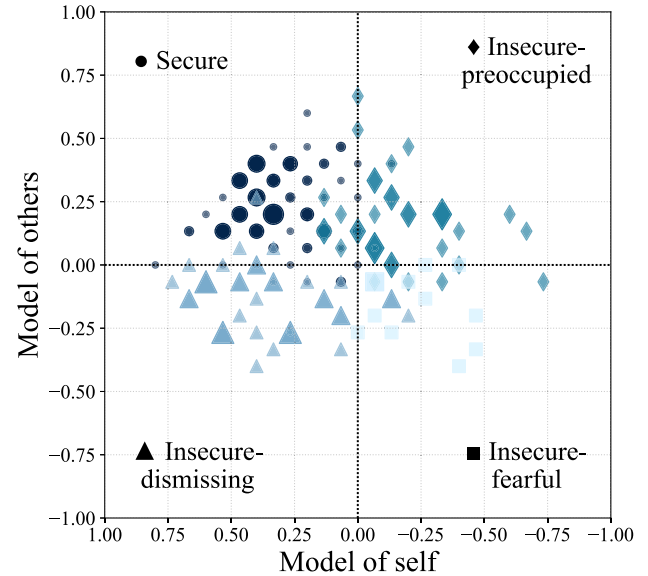


Fig. 3. Distribution of the attachment style labels based on the RQ as a function of the self and other-model scores for each subject. The size of the marker is proportional to the number of occurrences.

for the different types of classifiers trained, and the different questions used as input to the models. The figure presents the mean and standard deviation of accuracy, obtained in stratified 5-fold cross-validation for the models trained for each open-ended question independently and in group 5-fold when the concatenation of all of them was used. Notably, the best performing models in cross-validation were SVM and LR in both gender-dependent and gender-independent approaches, for all the questions considered. Therefore, these were the two classifiers selected for testing.

Table V shows the best classification results corresponding to the three different database configurations. The table includes the open-ended question from which the acoustic features were extracted (Question), the machine learning model used among the four tested (SVM, LR, RF and KNN), the number of selected features ( $N$ ) to create the corresponding model. Different evaluation metrics have been included in the table for cross-validation and testing: CV accuracy, CV AUC, test accuracy, test AUC, test sensitivity and test specificity. Moreover, the significant results for the permutation tests are indicated with an asterisk in the test accuracy column.

Regarding the models created with the complete database, referred to henceforth as the gender-independent models, the highest cross-validation accuracy, 93.65%, was achieved with a LR model using 88 features extracted from all open-ended questions, but the accuracy dropped to 52.31% in test. When

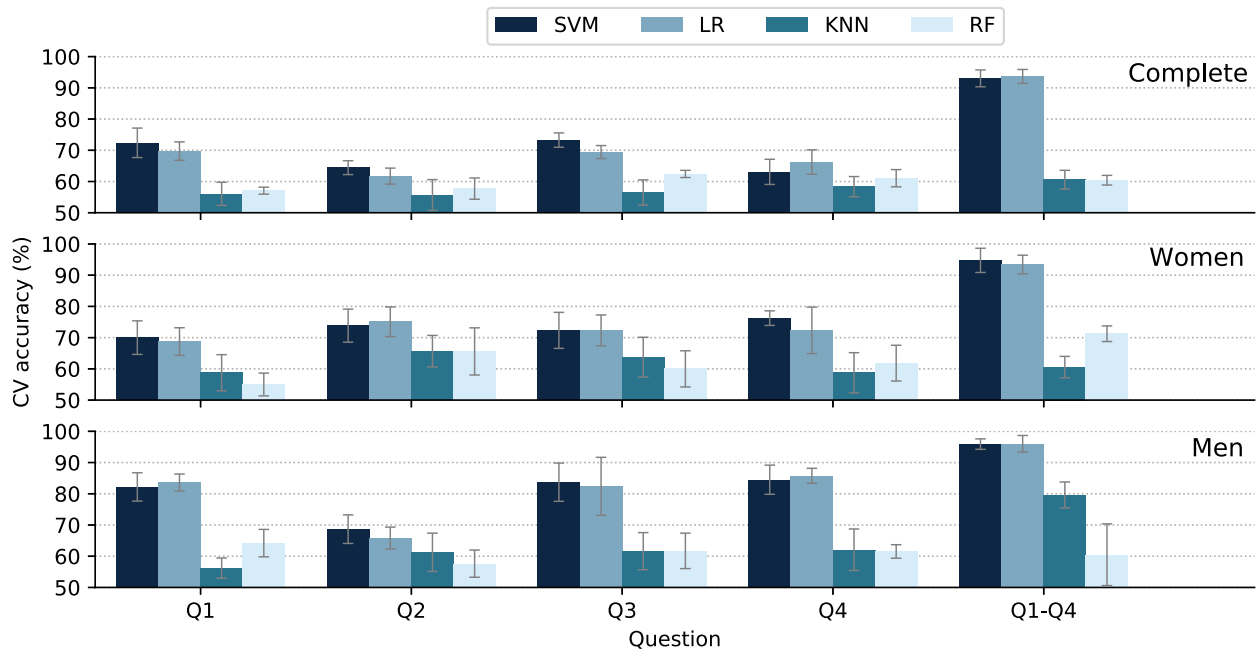


Fig. 4. Comparison of the cross-validation results for the different datasets, type of questions and classifiers.

TABLE V  
BEST CLASSIFICATION RESULTS FOR THE DIFFERENT DATASETS, TYPE OF QUESTIONS AND THEIR COMBINATION

Dataset	Question	Model	N	CV acc (%)	CV AUC	Test acc (%)	Test AUC	Test sensitivity	Test specificity
Complete	Q1	SVM	34	72.39 (9.44)	0.720 (0.094)	43.06 (8.51)	0.431 (0.083)	0.369 (0.120)	0.494 (0.133)
	Q2	SVM	72	64.44 (4.44)	0.615 (0.041)	39.50 (7.67)	0.394 (0.080)	0.230 (0.103)	0.559 (0.133)
	Q3	SVM	26	<b>73.27 (4.59)</b>	<b>0.727 (0.054)</b>	<b>58.88 (9.11)*</b>	<b>0.588 (0.093)</b>	<b>0.555 (0.130)</b>	<b>0.620 (0.124)</b>
	Q4	LR	12	66.23 (7.83)	0.646 (0.075)	56.87 (8.64)*	0.571 (0.086)	0.494 (0.138)	0.647 (0.112)
	Q1-Q4	LR	88	93.65 (4.46)	0.933 (0.047)	52.31 (8.31)	0.521 (0.086)	0.550 (0.133)	0.492 (0.110)
Women	Q1	SVM	26	70.00 (10.75)	0.701 (0.110)	50.38 (12.21)	0.495 (0.124)	0.504 (0.154)	0.487 (0.195)
	Q2	LR	22	75.08 (9.49)	0.757 (0.093)	57.75 (12.78)*	0.579 (0.133)	0.653 (0.174)	0.505 (0.214)
	Q3	SVM	10	<b>72.33 (11.53)</b>	<b>0.719 (0.122)</b>	<b>63.88 (11.48)*</b>	<b>0.639 (0.113)</b>	<b>0.767 (0.138)</b>	<b>0.510 (0.166)</b>
	Q4	SVM	36	76.25 (4.68)	0.762 (0.049)	58.00 (15.71)*	0.581 (0.118)	0.887 (0.105)	0.274 (0.185)
	Q1-Q4	SVM	68	94.75 (7.76)	0.945 (0.080)	31.00 (11.66)	0.313 (0.118)	0.373 (0.200)	0.252 (0.161)
Men	Q1	LR	34	83.62 (5.41)	0.826 (0.070)	56.13 (13.55)	0.563 (0.136)	0.632 (0.191)	0.493 (0.177)
	Q2	SVM	52	68.67 (9.16)	0.678 (0.082)	50.25 (10.82)	0.504 (0.111)	0.480 (0.167)	0.527 (0.198)
	Q3	SVM	32	<b>83.71 (12.26)</b>	<b>0.824 (0.118)</b>	<b>69.38 (12.58)*</b>	<b>0.691 (0.131)</b>	<b>0.623 (0.171)</b>	<b>0.758 (0.173)</b>
	Q4	LR	18	85.75 (4.79)	0.851 (0.049)	44.63 (11.92)	0.446 (0.123)	0.256 (0.158)	0.636 (0.168)
	Q1-Q4	LR	34	<b>96.00 (5.33)</b>	<b>0.956 (0.057)</b>	<b>83.63 (8.65)*</b>	<b>0.839 (0.086)</b>	<b>0.890 (0.118)</b>	<b>0.788 (0.152)</b>

\*  $p < 0.05$  in the t-test comparing with the accuracy distribution of the permutation test

Results with the highest test accuracy scores are marked in bold for each question, as well as grouped questions, if significant.

considering the open-ended questions independently, the models achieved a CV accuracy from 70.51% to 73.27%. Namely, the SVM classifier with 26 acoustic features from Q3 achieved the highest CV accuracy and a test accuracy of 58.88%.

The women and men models, referred to hereafter as the gender-dependent models, outperformed the gender-independent models. On the one hand, for the women dataset, the model trained using the concatenation of acoustic features extracted from the four open-ended questions achieved the highest CV scores, 94.75% accuracy, but again the test accuracy dropped to random values. With regard to the models trained with each open-ended question, they got CV accuracies from 73.67% to 76.25%, but lower values were obtained for testing. The SVM trained with 10 features from Q3 reached 73.67% in cross-validation and the highest test accuracy of 63.88%. On the other hand, the men dataset achieved the highest CV

accuracies for all the feature combinations. 34 features from all the open-ended questions were used to train the LR that reached 96% and 83.63% accuracy in CV and test respectively. As for the models trained for each particular question, the CV accuracies ranged from 72.83% to 85.75%. The SVM trained with 32 features from Q3 achieved 83.71% CV accuracy and 69.38% accuracy in test.

Turning now to the relevant features for attachment recognition, Table VI shows the selected features for each dataset's best model, which corresponds to Q3 in all cases plus the model based on all questions for the men dataset. In the latter, it has been detailed to which question each chosen characteristic corresponds. Moreover, the total number of features per category has been indicated in brackets in the Features column. For the gender-independent model based on Q3, twelve out of twenty-six features pertained to the spectral domain, specifically



TABLE VI  
SELECTED FEATURES FOR EACH OF THE BEST CLASSIFICATION MODELS

Category	Features	Complete Q3	Women Q3	Men Q3	Men Q1 - Q4
Frequency	Jitter (2)	1	0	1	1 (Q2), 1 (Q4)
	F0 (10)	1	0	4	3 (Q1), 1 (Q3), 1 (Q4)
	F1, F2, F3 (frequency) (6)	1	1	2	1 (Q2), 1 (Q3)
	F1, F2, F3 (bandwidth) (6)	2	1	3	1 (Q1), 1 (Q3)
Energy / amplitude	Shimmer (2)	0	0	2	2 (Q2)
	Loudness (10)	5	2	4	1 (Q1), 1 (Q3), 2 (Q4)
	HNR (2)	1	0	0	1 (Q2)
Spectral	Alpha Ratio (3)	1	0	0	1 (Q2)
	Hammarberg Index (3)	0	0	2	2 (Q1), 1 (Q2)
	Spectral slope (6)	1	0	2	1 (Q1), 1 (Q2)
	F1, F2, F3 (rel. energy) (6)	2	4	1	1 (Q1)
	H1-H2 (2)	0	0	1	0
	H1-A3 (2)	0	0	0	0
	1-4 MFCC (16)	7	2	6	1 (Q1), 1 (Q2), 2 (Q3), 2 (Q4)
	Spectral flux (5)	1	0	0	0
Temporal	Rate loudness peaks (1)	0	0	0	0
	Length voiced regions (2)	1	0	1	1 (Q2)
	Length unvoiced regions (2)	0	0	1	1 (Q1), 1 (Q2)
	Pseudo Syllable Rate (1)	1	0	1	0
Other	Equivalent sound level (1)	1	0	1	0
	Duration (1)	0	0	0	0
-	Total	26	10	32	34

seven were functionals of the MFCCs. Moreover, five of them were related to loudness and another five to formant frequency. With regard to the gender dependent models, the model trained with the women partition used only ten features, six of them from the spectral domain, two from the frequency domain and another two from the energy domain. For the Q3-based men model, thirty-two features were used: ten frequency features, six energy features, twelve spectral features, three temporal features and the equivalent sound level feature. Finally, regarding the men model created using features from Q1 to Q4, thirty-four acoustic features were used: eleven from Q1, eleven from Q2, six from Q3 and six from Q4. Most of them belonged to the spectral domain, with thirteen features selected from this category, as well as to the frequency domain, with eleven features used in the model.

## VI. DISCUSSION

This section is divided into four subsections. Firstly, main results are presented. Secondly, a comparison with previous research is conducted. Next, limitations and future research are discussed. Finally, the study implications are described.

### A. Principal Results

We created a secure and insecure attachment recognition system based on acoustic features extracted from online recordings of individuals' natural responses to four open-ended questions intended to trigger their attachment system. We investigated the impact of gender on the recognition system by developing models that were either gender-independent or gender-dependent. We gathered responses from 96 women and 93 men, and we used a set of standard acoustic features to train different machine learning models.

The gender-dependent models outperformed the gender-independent approach when considering each question independently. The models based on the complete dataset achieved CV accuracies of 64.44% to 73.27%, and the best results in test was 58.88% accuracy using features from question Q3. Conversely,

gender-dependent models reached CV accuracies of 68.67% to 85.75%, and test accuracies higher than 60%. Specifically, the women models got CV accuracy values between 70.00% and 76.25% and the highest performance in test was 63.88% accuracy using Q3. The men models had better performance, up to 85.75% CV accuracy, and the highest results in test was 69.38% accuracy, 0.623 sensitivity and 0.758 specificity, again with acoustic features from Q3. Out of the four open-ended questions, this particular one was intended to activate the attachment system by inquiring about the social context. It directly asked participants about their interpersonal relationships, including their present status and if they would want to change anything about them.

With regard to models created using the combination of features from all the questions, the gender-dependent classifiers again outperformed the gender-independent models. The CV accuracy for the model using the complete database was 93.65% and 52.31% in test, whereas for women it was 94.75% and 31% and for men it was 96% and 83.63% for CV and test accuracy respectively. Therefore, gender-dependent models achieved slightly better scores in CV. However only the men model got remarkable results in test, whereas the performance in the complete and women models indicates that they could not correctly generalize despite feature selection.

The superior performance of the gender-dependent models may indicate the existence of differences in the way attachment is manifested through voice between both genders. Accordingly, Zhang and Zheng [31] found that vocal cues related to fundamental frequency were correlated with avoidant attachment style in men, whereas no correlations were found for women. Moreover, consideration of gender has proven to improve the classification results in speech-based depression detection [41] and speech emotion recognition [42]. Nevertheless, these findings could also be the result of other two potential factors. On the one hand, substantial differences in gender-specific speech characteristics could prevent a unified model from discerning patterns. This may occur especially in cases where there is

little data, as in this study, making it difficult for the model to generalize. On the other hand, it may be the case that current data modeling architectures lack the capacity to capture gender-independent patterns effectively. However, to ensure comparability of outcomes between the gender-dependent and gender-independent models, we used the same data modeling pipeline for both approaches.

Regarding the features that were relevant for attachment classification, detailed in Table VI, the ones selected from Q3 for the three data partitions explored have some similarities. For the gender-independent model, loudness and MFCCs represented half of the features selected. Speech loudness can change due to alterations in the vocal effort and tract resonance, both of which may be influenced by the physiological effects of emotions [43]. MFCCs have proven to be useful for depression detection [44] and emotion recognition [45] from speech, which involves similar implicit processes to those hypothesized to occur in attachment behaviors. However, the low test scores of this model allow no meaningful conclusions to be drawn. As for the gender-dependent models, the women model approach showed a similar trend: the selected features were related to loudness, MFCCs and also F1-F3 formants. Formants are associated with vowels and voiced consonants in spoken language and, although speech production is largely under voluntary control, it is still susceptible to involuntary perturbations [43]. The formants amplitudes and bandwidths can be affected by the amount of saliva in the mouth, between other factors, which can decrease in some speakers in certain situations such as public presentations. In this investigation, it could have occurred in insecure participants due to anxiety about answering the question. With regard to the men model, which performed best in test among the three approaches, it differed in the selected features from the women model. Temporal features including the length of the voiced and unvoiced regions and the pseudo syllable rate were considered relevant for classification. These temporal features may be related to the nervousness or anxiety experienced by the subjects when answering the questions. F0 based features including jitter, shimmer and functionals of F0 were also selected, and they have shown to be altered by anxiety, stress and emotions [46], [47]. MFCCs, loudness and F1-F3 features were also included, similarly to the women model. Considering the men model created using acoustic features from questions Q1-Q4, which outperformed the previous approaches with a test accuracy of 83.63%, it used characteristics uniformly distributed among the different questions and categories. Similarly to previous results, MFCCs features were selected from all the questions, as well as formants and loudness. The length of voiced and unvoiced regions from Q1 and Q2 was also selected, as well as voice quality features (jitter, shimmer, HNR). It is noteworthy that in no combination was the audio duration selected as a relevant feature for attachment style classification.

The methodology proposed in this research aimed to trigger the attachment system with an interrogation system that included four specific open-ended questions to be answered online. Subjects were also required to respond the questions while alone in a room to further facilitate the attachment system activation. The high recognition rates reached with the gender-dependent models, specially with the men model, show that the questions

had an impact on the way attachment style was manifested through voice. By utilizing a machine learning framework with a multivariate approach, this study was able to effectively differentiate between secure and insecure attachment styles and develop a recognition model for future subjects. This sets it apart from classical methods like statistical inference, which focuses on identifying relationships between variables rather than group discrimination. However, machine learning models have some limitations that need to be considered. They sacrifice interpretability for predictive power, so they are sometimes considered as “black boxes”. To overcome this limitation, an automatic feature selection wrapper was utilized in this study to identify and analyze the selected features in the best models. In addition, the limited size of the data may also limit the generalizability of the models. Nevertheless, a test set of 32 participants was included here to verify that the cross-validation metrics of the models do not suffer from overfitting.

## B. Comparison With Previous Research

Previous studies have analyzed the relationship between voice and attachment style using samples involving between 10 and 76 subjects [25], [26], [27], [28], [29], [30], and exceptionally 206 participants in [31] and 206 (finally used 118) participants in [32]. We recruited 199 participants for our analysis: this is more than double the number of participants used in the majority of investigations and matches the only two studies with a similar sample size [31], [32]. Concerning the methodology, we implemented a web-based interrogation system that included reflective questions designed to trigger the attachment system, which could have facilitated the ability to detect the manifestation of attachment style in the voice; this is in contrast to other investigations [25], [26], [31] that did not consider this factor. The questions were also self-administered, so they could be answered at home in a familiar environment and without the need for physical displacement, overcoming limitations of previous studies in laboratory conditions [25], [26], [27], [28], [29], [30], [31], [32]. In terms of the analysis, only two previous studies developed machine learning models for attachment recognition. Parra et al. [28] used multimodal features collected during the Biometric Attachment Test to predict attachment style. Paralinguistic features extracted from speech proved crucial for avoidance attachment detection and less significant for anxiety detection. Similarly, our results showed that features extracted from speech had enough information to distinguish between secure and insecure adults, and our gender-balanced sample allowed us to find better performance for gender-dependent models. Koçak et al. [32] developed machine learning models to predict attachment styles obtaining good performances, but certain differences with respect to our study are worth noting. On the one hand, they used only secure and fearful attachment styles, which are the most extreme cases and probably the easiest to differentiate. In contrast, in our study we used all subtypes of insecure attachment style, which results in a more heterogeneous data under the insecure category, as shown in Fig. 3, but also represents the real distribution of attachment styles. Moreover, they evaluated the models using leave-one-out

cross validation, and the same strategy was used for the hyperparameter tuning, so their results may be slightly optimistic as they also claim. Conversely, we used a cross-validation strategy to optimize the hyperparameters and a separate test partition to finally evaluate the models in order to avoid optimistic results and to assess possible overfitting. On the other hand, they got higher performances by combining both genders rather than considering them independently. However, they do not filter the data based on gender, but include spouse's features when predicting their partner's attachment style instead of considering only their own features. Therefore, this consideration of gender is not comparable with our approach, since we used the same characteristics for both genders and the difference was in the data used for training the models and not in the features. In addition, other studies in the literature have explored the relationship between attachment style and the voice using statistical analysis. However, our findings are not directly comparable, as the speech samples were collected under different assumptions: [25], [26], [31] used speech samples not representative of usual speech, and in [29], [30] the analysis was performed at word level. However, it is noteworthy that we obtained the best classification results for the men model, and likewise in [31] correlations between attachment style and vocal cues were only found in men.

### C. Limitations and Future Research

Certain limitations of this research need to be considered. First, the attachment label was obtained from a self-report measure, which is criticized for its low ecological validity and social desirability bias. However, we decided to administer the Relationship Questionnaire because of its shortness, a crucial factor when the sample is collected through online panels. In the future, it would be necessary to combine behavioral measures with expert assessment to overcome the limitations of considering questionnaires as the gold-standard. Unsupervised learning would also be a powerful alternative, as this machine learning approach could determine patterns from data without the labels. However, even taking into account the ground truth used, applying the general biomarker patterns from a large number of subjects as a reference for online attachment assessment reduces the likelihood of misrepresentation in responses, thereby increasing the objectivity of the assessment system. Another limitation of the present study was the inability to control all factors that may influence the data collected with online systems, such as interruptions during the research or the audios' forwarding. In fact, we decided to exclude audio samples considered as noisy, which could limit the results of the model in real-world conditions. Nonetheless, it must be noted that online systems offer natural and comfortable conditions for the user that can overcome other limitations associated with laboratory environments and may also facilitate attachment system activation. It is worth mentioning that the study sample had a slight underrepresentation of older individuals, with those aged 60 years or older being excluded from the study. Given that older adults tend to have a more consistent attachment style, as mentioned in the introduction, it would be valuable to gather data from this age group and create an attachment recognition model to determine whether stability factors may potentially

lead to improved classification outcomes. Further studies are needed to investigate the effect of study design parameters, such as the order of the questions or the time between them, and the impact of language in attachment recognition to ascertain whether the same recognition models can be applied to different languages. New model architectures could also be explored to improve the results of the gender-independent model. Finally, it would be interesting to further investigate the differentiation of subtypes within insecure attachment by developing a four-class classification model, which would require the use of a carefully balanced database across the four attachment styles.

### D. Implications

Attachment recognition has the potential to be applied to a wide variety of fields, particularly healthcare, as insecure attachment styles have implications for people's physical and mental health. Moreover, the speech activities used in this research were designed to be completed online, using a web-based interrogation system on a mobile phone or computer. Participants could thus complete the study at home, in comfortable conditions, overcoming the limitations of the laboratory setting. Therefore, our methodology could be used to recognise insecure attachment using a remote system, such as a mobile phone, which would allow this health risk factor to be identified quickly and objectively, supporting large-scale mobile screening systems. In addition, this study included four open-ended questions that were automatically presented to the individuals via text on the screen. This constitutes a first approach to a potential intelligent interrogation system that could be used for online assessments in several psychological domains such as personality dimensions or emotional and/or cognitive clinical conditions. Lastly, our findings suggest that gender differences exist in the way an attachment system is manifested through the voice, as higher results are achieved when considering gender-dependent models. Consequently, further research should consider both genders separately.

## VII. CONCLUSION

This study developed attachment recognition models for men and women independently, exclusively based on acoustic features extracted from speech recordings obtained with a web-based interrogation system. Several limitations in the literature were overcome. Data collection included samples of 199 subjects from the general population balanced in terms of gender and secure and insecure attachment style. The methodology was designed to trigger the attachment system and obtain speech samples that were representative of typical speech. By using an independent test partition, a classification accuracy of 63.88% in the women model and 83.63% in the men model was achieved, showing gender differences in how attachment influences the voice. The procedure was designed to be easily applied remotely in a large sample, facilitating the potential application of attachment recognition in clinical and research environments. However, the limitations included the reliance on self-report measures, the underrepresentation of older individuals, and the exclusion of noisy recordings. Future research should combine self-report measures with expert assessments to

obtain the labeling, gather data on elderly individuals, include noise removal techniques and investigate the impact of study parameters such as question order to enhance classification outcomes and overcome potential biases.

## REFERENCES

- [1] J. Bowlby, *Attachment and Loss* (Attachment), vol. 1. New York, NY, USA: Basic Books, 1969.
- [2] J. Bowlby, *Attachment and Loss* (Separation: Anxiety and Anger), vol. 2. New York, NY, USA: Basic Books, 1973.
- [3] J. Bowlby, *Attachment and Loss* (Loss: Sadness and Depression), vol. 3. New York, NY, USA: Basic Books, 1980.
- [4] K. Bartholomew, "Avoidance of intimacy: An attachment perspective," *J. Soc. Pers. Relationships*, vol. 7, no. 2, pp. 147–178, 1990.
- [5] K. Bartholomew and L. M. Horowitz, "Attachment styles among young adults: A test of a four-category model," *J. Pers. Social Psychol.*, vol. 61, no. 2, pp. 226–244, 1991.
- [6] M. Mikulincer and P. R. Shaver, "The attachment behavioral system in adulthood: Activation, psychodynamics, and interpersonal processes," in *Advances in Experimental Social Psychology*. San Diego, CA, USA: Elsevier, 2003, pp. 53–152.
- [7] Z. Mortazavizadeh and S. Forstmeier, "The role of emotion regulation in the association of adult attachment and mental health: A systematic review," *Arch. Psychol.*, vol. 2, no. 9, pp. 1–25, 2018.
- [8] J. A. Feeney, "Implications of attachment style for patterns of health and illness," *Child: Care, Health Develop.*, vol. 26, no. 4, pp. 277–288, 2000.
- [9] L. A. McWilliams and S. J. Bailey, "Associations between adult attachment ratings and health conditions: Evidence from the National Comorbidity Survey Replication," *Health Psychol.*, vol. 29, no. 4, pp. 446–453, 2010.
- [10] E. T. Huntsinger and L. J. Luecken, "Attachment relationships and health behavior: The mediational role of self-esteem," *Psychol. Health*, vol. 19, no. 4, pp. 515–526, 2004.
- [11] M. Del. Giudice, "Sex differences in attachment styles," *Curr. Opin. Psychol.*, vol. 25, pp. 1–5, 2019.
- [12] M. Del. Giudice, "Sex-biased ratio of avoidant/ambivalent attachment in middle childhood," *Brit. J. Develop. Psychol.*, vol. 26, pp. 369–379, 2008.
- [13] G. Gloger-Tippelt and G. Kappler, "Narratives of attachment in middle childhood: Do gender, age, and risk-status matter for the quality of attachment?," *Attachment Hum. Develop.*, vol. 18, no. 6, pp. 570–595, Nov. 2016.
- [14] M. D. Giudice and R. Angeleri, "Digit ratio (2D:4D) and attachment styles in middle childhood: Indirect evidence for an organizational effect of sex hormones," *Adaptive Hum. Behav. Physiol.*, vol. 2, pp. 1–10, 2016.
- [15] M. D. Giudice, "Sex differences in romantic attachment: A meta-analysis," *Pers. Social Psychol. Bull.*, vol. 37, no. 2, pp. 193–214, Jan. 2011.
- [16] D. P. Schmitt, "Are men universally more dismissing than women? Gender differences in romantic attachment across 62 cultural regions," *Pers. Relationships*, vol. 10, no. 3, pp. 307–331, Sep. 2003.
- [17] M. D. Giudice, "Sex differences in romantic attachment: A facet-level analysis," *Pers. Individual Differences*, vol. 88, pp. 125–128, 2016.
- [18] P. Ravitz, R. Maunder, J. Hunter, B. Sthankiya, and W. Lancee, "Adult attachment measures: A 25-year review," *J. Psychosomatic Res.*, vol. 69, no. 4, pp. 419–432, 2010.
- [19] B. A. Nosek, C. B. Hawkins, and R. S. Frazier, "Implicit social cognition: From measures to mechanisms," *Trends Cogn. Sci.*, vol. 15, no. 4, pp. 152–159, Apr. 2011.
- [20] P. S. Forscher et al., "A meta-analysis of procedures to change implicit measures," *J. Pers. Social Psychol.*, vol. 117, pp. 522–559, 2019.
- [21] C. George, N. Kaplan, and M. Main, "The adult attachment interview," 1985.
- [22] D. M. Rosenthal, "Intentionality," *Midwest Stud. Philosophy*, vol. 10, pp. 151–184, 1986.
- [23] M. B. Akçay and K. Oğuz, "Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers," *Speech Commun.*, vol. 116, pp. 56–76, 2020.
- [24] A. S. Cohen and B. Elvevåg, "Automated computerized analysis of speech in psychiatric disorders," *Curr. Opin. Psychiatry*, vol. 27, pp. 203–209, 2014.
- [25] E. Monti, D. C. Kidd, L. M. Carroll, and E. Castano, "What's in a singer's voice: The effect of attachment, emotions and trauma," *Logopedics Phoniatrics Vocol.*, vol. 42, pp. 62–72, 2016.
- [26] D. R. Becker, "Vocal manifestations of reported past trauma." Jun. 2019, to be published. [Online]. Available: <http://d-scholarship.pitt.edu/36599/>
- [27] F. Parra, R. Miljkovitch, G. Persiaux, M. Morales, and S. Scherer, "The multimodal assessment of adult attachment security: Developing the biometric attachment test," *J. Med. Internet Res.*, vol. 19, no. 4, Apr. 2017, Art. no. e100.
- [28] F. Parra, S. Scherer, Y. Benezeth, P. Tsvetanova, and S. Tereno, "Development and cross-cultural evaluation of a scoring algorithm for the biometric attachment test: Overcoming the challenges of multimodal fusion with "small data,"" *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, pp. 211–225, Jan.–Mar. 2022.
- [29] M. Spinelli, M. Fasolo, G. Coppola, and T. Aureli, "It is a matter of how you say it: Verbal content and prosody matching as an index of emotion regulation strategies during the adult attachment interview," *Int. J. Psychol.*, vol. 54, pp. 102–107, 2019.
- [30] M. Moneta, M. Penna, H. Loyola, A. Buchheim, and H. Kächele, "Measuring emotion in the voice during psychotherapy interventions: A pilot study," *Biol. Res.*, vol. 41, pp. 389–395, Feb. 2008.
- [31] J. Zhang and L. Zheng, "Masculine voices predict attachment style and relationship communication patterns in romantic relationships," *J. Sex Marital Ther.*, vol. 47, no. 3, pp. 262–269, 2021.
- [32] T. M. Koçak, B. Ç. Dibek, E. N. Polat, N. Kafesçioğlu, and C. Demiroğlu, "Automatic detection of attachment style in married couples through conversation analysis," *EURASIP J. Audio, Speech, Music Process.*, vol. 2023, no. 1, 2023, Art. no. 26.
- [33] M. Mikulincer, O. Gillath, and P. Shaver, "Activation of the attachment system in adulthood: Threat-related primes increase the accessibility of mental representations of attachment figures," *J. Pers. Social Psychol.*, vol. 834, pp. 881–895, 2002.
- [34] S. Grüner, N. Hirschauer, and O. Mußhoff, "The potential of different experimental designs for policy impact assessment," *German J. Agricultural Econ.*, vol. 65, no. 03, pp. 159–170, 2016.
- [35] F. Eyben et al., "The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing," *IEEE Trans. Affect. Comput.*, vol. 7, no. 2, pp. 190–202, Apr.–Jun. 2015.
- [36] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: The munich versatile and fast open-source audio feature extractor," in *Proc. 18th ACM Int. Conf. Multimedia*, 2010, pp. 1459–1462.
- [37] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. IEEE 8th Int. Conf. Data Mining*, 2008, pp. 413–422.
- [38] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, pp. 389–422, 2002.
- [39] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, 2020.
- [40] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *Int. J. Data Mining Knowl. Manage. Process.*, vol. 5, no. 2, pp. 1–11, 2015.
- [41] B. Vlasenko, H. Sagha, N. Cummins, and B. Schuller, "Implementing gender-dependent vowel-level analysis for boosting speech-based depression recognition," in *Proc. InterSpeech*, 2017, pp. 3266–3270.
- [42] I. Bisio, A. Delfino, F. Lavagetto, M. Marchese, and A. Sciarone, "Gender-driven emotion recognition through speech signals for ambient intelligence applications," *IEEE Trans. Emerg. Topics Comput.*, vol. 1, no. 2, pp. 244–257, Dec. 2013.
- [43] T. Johnstone, "The effect of emotion on voice production and speech acoustics," Ph.D. dissertation, Psychology Department, Univ. Western Australia, Crawley WA, Australia, 2017.
- [44] T. Taguchi et al., "Major depressive disorder discrimination using vocal acoustic features," *J. Affect. Disord.*, vol. 225, pp. 214–220, 2018.
- [45] A. Koduru, H. B. Valiveti, and A. K. Budati, "Feature extraction algorithms to improve the speech emotion recognition rate," *Int. J. Speech Technol.*, vol. 23, no. 1, pp. 45–55, 2020.
- [46] B. F. Fuller, Y. Horii, and D. A. Conner, "Validity and reliability of nonverbal voice measures as indicators of stressor-provoked anxiety," *Res. Nurs. Health*, vol. 15, no. 5, pp. 379–389, 1992.
- [47] X. Li et al., "Stress and emotion classification using jitter and shimmer features," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2007, vol. 4, pp. IV-1081–IV-1084.