Measuring Stress in Health Professionals over the Phone using Automatic Speech Analysis during COVID-19 Pandemic: Observational Study

Alexandra König, Kevin Riviere, Nicklas Linz, Julia Elbaum, Roxane Fabre, Philippe Robert

Submitted to: Journal of Medical Internet Research
on: September 28, 2020

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on it's website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressively prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Measuring Stress in Health Professionals over the Phone using Automatic Speech Analysis during COVID-19 Pandemic: Observational Study

Alexandra König[1] BSc, PhD, MSc; Kevin Riviere[2]; Nicklas Linz[3]; Julia Elbaum[2]; Roxane Fabre[2]; Philippe Robert[4]

[1]Institut national de recherche en informatique et en automatique (INRIA), Stars Team, Sophia Antipolis, France Valbonne FR
[2]Département de Santé Publique, Centre Hospitalier Universitaire de Nice, Université Côte d'Azur, Nice, France Nice FR
[3]ki elements Saarbrücken DE
[4]CoBteK (Cognition-Behaviour-Technology) Lab, FRIS - University Côte d'azur, Nice, France Nice FR

Corresponding Author:
Alexandra König BSc, PhD, MSc
Institut national de recherche en informatique et en automatique (INRIA), Stars Team, Sophia Antipolis, France
2004 Route des Lucioles, 06902 , Sophia Antipolis, France
Valbonne
FR

Abstract

Background: During the current COVID-19 pandemic, health professionals are directly confronted with the suffering of patients and their families. By making them main actors in the management of this health crisis, they are exposed to various psychosocial risks (stress, trauma, fatigue, etc.). Paradoxically, stress-related symptoms are often underreported in this vulnerable population but potentially detectable through passive monitoring of changes in speech behavior.

Objective: The study aims to investigate the use of a rapid and remote measure of stress levels in health professionals working during this COVID 19 outbreak through the analysis of their speech behavior during a short phone call conversation, and in particular a positive/negative and neutral story telling task.

Methods: For this, speech samples of 89 healthcare professionals were collected over the phone and various voice features extracted and compared with classical stress measures via standard questionnaires. Regression analysis was additionally performed.

Results: Certain speech characteristics correlated with stress levels in both genders; mainly spectral (formant) features as the Mel-frequency cepstral coefficients (MFCC) and prosodic characteristics such as the fundamental frequency (F0) seemed sensitive to stress. Overall, for both male and female participants, using vocal features from the positive tasks for regression yielded most accurate prediction results of stress scores (MAE = 5.31).

Conclusions: Automatic speech analysis could help with early detection of subtle signs of stress in vulnerable populations over the phone. Combining the use of this technology with timely intervention strategies it could contribute to the prevention of burn outs as well as the development of co-morbidities such as depression or anxiety.

(JMIR Preprints 28/09/2020:24191)
DOI: https://doi.org/10.2196/preprints.24191

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Measuring Stress in Health Professionals over the Phone using Automatic Speech Analysis during COVID-19 Pandemic: Observational Study

A.König[1,3], K. Riviere [2,3], N. Linz [5], H. Lindsay,[6],, J. Elbaum[2,3], R. Fabre[2,3], A. Derreumaux[3], P.H. Robert [2, 3,4]

[1] Institut national de recherche en informatique et en automatique (INRIA), Stars Team,
Sophia Antipolis, France

[2] Département de Santé Publique, Centre Hospitalier Universitaire de Nice, Université Côte d'Azur, Nice, France

[3] CoBteK (Cognition-Behaviour-Technology) Lab, FRIS - University Côte d'azur, Nice, France

[4] Association IA, Nice, France

[5] ki elements, Saarbrücken, Germany

[6] German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Germany

Corresponding Author:
Alexandra König, PhD
Cobtek (Cognition-Behaviour-Technology) Lab
University Côte d'azur
Institut Claude Pompidou
10 rue Molière,
06100, Nice, France
Tel. : +33 492 034 760
Email : alexandra.kong@inria.fr

Abstract

Background: During the current COVID-19 pandemic, health professionals are directly confronted with the suffering of patients and their families. By making them main actors in the management of this health crisis, they are exposed to various psychosocial risks (stress, trauma, fatigue, etc.). Paradoxically, stress-related symptoms are often underreported in this vulnerable population but are potentially detectable through passive monitoring of changes in speech behavior.

Objectives: This study aims to investigate the use of rapid and remote measures of stress levels in health professionals working during the COVID-19 outbreak. This was done through the analysis of their speech behavior during a short phone call conversation, and in particular a positive, negative and neutral story telling task.

Methods: Speech samples of 89 healthcare professionals were collected over the phone during a positive, negative and neutral story telling task and various voice features were extracted and compared with classical stress measures via standard questionnaires. Additionally, a regression analysis was performed.

Results: Certain speech characteristics correlated with stress levels in both genders; mainly spectral (formant) features as the Mel-frequency cepstral coefficients (MFCC) and prosodic characteristics such as the fundamental frequency (F0) appear sensitive to stress. Overall, for both male and female participants, using vocal features from the positive tasks for regression yielded the most accurate prediction results of stress scores (MAE = 5.31).

Conclusions: Automatic speech analysis could help with early detection of subtle signs of stress in vulnerable populations over the phone. Combining the use of this technology with timely intervention strategies, it could contribute to the prevention of burnout as well as the development of co-morbidities such as depression or anxiety.

Keywords: Stress detection, speech, voice analysis, COVID19, phone monitoring, computer linguistic

Introduction

In December 2019, in the Chinese city of Wuhan, a new coronavirus pneumonia (called COVID-19 for coronavirus disease 2019) emerged. The pathogen involved is SARS-CoV-2 (for severe acute respiratory syndrome coronavirus 2). Here we will refer to the pathology as COVID-19. COVID-19 has spread very rapidly in China but also in many other countries [1]. On March 11, 2020, the World Health Organization declared that COVID-19 had become a pandemic [2].

According to previous studies on SARS or Ebola epidemics, the onset of a sudden and immediately fatal disease could put extraordinary pressure on health care professionals [3]. Increased workloads, physical exhaustion, inadequate personal equipment, nosocomial transmission and the need to make ethically difficult decisions about rationing care can have dramatic effects on their physical and mental well-being. Their resilience may be further compromised by isolation and loss of social support, risk or loss of friends and relatives, and radical, often worrying changes in working methods. Health care workers are therefore particularly vulnerable to mental health problems, including fear, anxiety, depression and insomnia [4-5]. Initial results estimate that 23% and 22% of healthcare workers experienced depression and anxiety respectively during the COVID-19 pandemic [6].

Paradoxically, healthcare workers do not tend to seek professional help and stress-related symptoms are often not immediately reported: *"burnout, stress and anxiety will have to wait "* - most of the time there will not even be a demand for care. Early implicit stress detection is of great importance in this population and would allow for timely intervention strategies in order to prevent escalation and complete occupational burnout.

To measure stress in clinical practice, various scales and questionnaires such as the Perceived Stress Scale (PSS) [7], Stressful Life Event Questionnaire [8] ; Stress Overload Scale [9] , or the Trier Inventory of Chronic Stress [10] are available. However, the present health crisis pushed research teams to investigate the use of new technological tools in this specific population. One possible avenue is the use of automatic speech analysis allowing extraction of voice features during standard consultation or over a simple phone call.

Psychological stress induces multiple effects on the body including increased muscle tension, breathing rate and changes in salivation rate that may in turn affect vocal production [11-12]. Under psychological stress, voice pitch (the acoustic correlate of fundamental frequency, F0) usually increases as it is inversely related to the rate of vocal fold vibration that stretch under stress and becomes tenser together with an increase in sub-glottal pressure and vocal intensity [13-14]. Indeed, an increase in voice pitch is the most commonly reported finding in studies examining speech under stress. Although, stress can also affect other voice parameters such as an increase of the speech prosody [11;13]. In depression, the analysis of speech characteristics has recently attracted considerable research attention [15-17]. Studies reveal that patients show flattened affect, reduced speech variability and monotonicity in pitch and loudness as well as increased pause duration and reduced speech rate [18-20]. A recent study investigated the use of speech parameters extracted from audio recordings to differentiate patients suffering from post-traumatic stress disorder from healthy controls [21].

Thus, the detection of subtle events in the voice may offer a window into assessing the impact of stress in situations where circumstances make it difficult to monitor stress directly but need to be addressed urgently [22].

In the present work, we aim to investigate the use of a rapid and remote measure of stress levels in health professionals working during the COVID-19 outbreak utilizing the automatic analysis of their speech behavior during a short phone call conversation.

For this, firstly, speech samples of healthcare professionals were collected over the phone during the COVID-19 pandemic and various voice features were extracted and compared with classical stress measures. Secondly, based on the extracted features participants' obtained scores on the completed stress scale were predicted. The purpose of this pilot study was to assess if this technological method could be of interest to support early screening of subtle signs of stress.

Methods

Participants

Healthcare professionals were recruited through outreach telephone calls. They worked during the COVID outbreak in the local University hospital center of Nice, in either private practices or as independent workers in the PACA (Provence-Alpes-Côte d'Azur) region. They could occupy any function in these structures. The only criterion for non-inclusion was the subjects' refusal to participate in the study. Inclusions were carried out from 05 May 2020 to 07 June 2020.

The study was approved by the Ethical Board for non-interventional studies of the University Côte d'azur, France (Approval No 2020-58). Participants were given all information about the study prior to the call so they could give informed consent. For those interested, the option for a follow-up call with a clinician was provided.

Procedure

The telephone calls were made by psychiatrists (n=3) or psychologists (n=1) belonging to the CoBTeK research team and the Memory clinic of the University Côte d'Azur. Calls lasted about fifteen minutes and were composed as follow:

a. An informative part explaining the reasons for the call, its structure, and how the study is conducted. The participant's consent will be requested to continue and to proceed with a recording of his or her voice.

b. MSA (motivation, stress, affect) questionnaire: The MSA questionnaire is a self-questionnaire composed of 11 questions which must be answered by yes or no. The first five questions assess motivation [23], the next two questions assess depression, and the last four questions assess stress [24].

c. Three open standardized questions (neutral, positive and negative story telling) : In order to capture natural speech but within a limited timeframe, the participant is asked to tell something emotionally neutral (describe where he or she is), to tell about a negative event in his or her life and finally to tell about a positive event in his or her life. Each answer should last about one minute and is recorded in a secure and encrypted way. It is not specified if the event has to be experienced during COVID, thus it is open to the participant to recall whatever event first comes to mind. These free speech tasks were used in previous studies [25; 18] and allow for greater range of induced emotional effects, potentially sensitive to signs of stress and depression. The comparison of speech features between neutral and emotionally loaded questions may give insight to the affective state of participants.

d. A perceived stress questionnaire: The Perceived Stress Questionnaire (PSQ) [26] is a hetero-questionnaire composed of 10 questions to be answered by "never", "almost never", "sometimes", "quite often", "often".

e. An open listening part aimed at exploring certain points in greater depth in order to refine the clinical needs.

f. Decision and advice: Following the above steps, the appellant will offer or not offer psychological follow-up if he or she considers that the patient is at risk of developing or has a mood or anxiety disorder. He or she may also offer advise on intervention strategies (relaxation, yoga, physical activity, national call platform for psychological support for caregivers).

Material

To perform the phone calls for this study, the application ‘DELTA’ phone version [27] (provided by https://ki-elements.de) was used. The DELTA solution allows the use of a dedicated interface in the form of an iOS application to make phone calls and locally record these calls on the internal memory of the iPad. The phone calls were made directly with the Ipad and through its internal microphone.
These recordings are then automatically transmitted (the iPad must be connected to the Internet) to the DELTA API for analysis of acoustic and semantic parameters. Once the analysis is completed, the results are displayed directly on the DELTA interface. The recordings are made locally on the phone, the connection between the interface and the DELTA API is secure and encrypted, and the recordings are destroyed from the DELTA servers once the analysis is completed and the results sent to the experimenter.

Analysis

Audio features were extracted directly and automatically from the recorded audio signals of the three open standardized questions (section c). Characteristics were extracted from four main areas:

- *Prosodic characteristics*, on long-term variations in perceived stress and speech rhythm. Prosodic features also measure alterations in per-sonic speech style (e.g., perceived pitch, speech intonation);
- *Formant characteristics* represent the dominant components of the speech spectrum and convey information about the acoustic resonance of the vocal tract and its use. These markers are often indicative of articulatory coordination problems in motor speech control disorders;

- *Source characteristics* that are related to the source of voice production, the airflow through the glottal speech production system. These features make operational irregularities in the movement of the vocal fold (e.g. voice quality measurements).
- *Temporal characteristics* include measures of the proportion of speech (e.g. duration of pauses, duration of speech segments), speech segment connectivity, and overall speech rate.

Features were extracted using Python 3.7 [28] and free and publicly available packages. For the temporal features, the My Voice Analysis [29] package was used. This package is built off of the speech analysis research tool praat [30]. Temporal features are actualized as the speech rate, syllable count, rate of articulation, speaking duration, total duration and ratio of speaking to non-speaking. This package is also used to extract prosodic features, namely the fundamental frequency values (F0; mean, standard deviation, minimum, maximum, upper and lower quartile). The F0 value is the representation of what is known as the pitch.

Formant features are calculated using the Python Speech Features library [31]. To characterize this aspect of speech, the original sound recording is refit according to a series of transformations commonly used for speech recognition that yield a better representation of the sound called the Mel Frequency Cepstrum (MFC). From this new representation of the sound form, the first fourteen coefficients of the MFC are extracted. The MFC values are extracted given that they describe the spectral shape of the audio file, generally with diminishing returns in terms of how informative they are, which is why we only consider the first 14 coefficients. If we were to select a greater number of MFC values it would result in a potentially needlessly more complex machine learning model using less informative features.

From each of these waves, the mean, variance, skewness and kurtosis are calculated for the energy (static coefficient), velocity (first differential), and acceleration (second differential).

The Librosa package [32] is used to calculate the mean, maximum, minimum and standard deviation of the root mean square value (RMS), centroid, bandwidth, flatness, zero crossing rate (ZCR), flatness, loudness, and flux of the spectogram, or the visualisation of the recording.

The source characteristics are extracted using the Signal Analysis package (Signal-Analysis==0.1.26) to extract the micro-movements of the sound wave; harmonic-to-noise ratio (HNR), Jitter, Shimmer and glottal pulses. Jitter and shimmer are two features of vocal signals that describe the frequency variation from cycle to cycle of the sound wave and the waveform amplitude, respectively [33,34]. While jitter rises with the growing lack of control for vocal cords vibration, higher shimmer is coupled with increased breathiness. HNR is the ratio between periodic components and non periodic components that constitute a voiced speech segment [35]. These components correspond to the vibration from vocal cords and glottal noise, respectively.

Speech features vary naturally between males and females due to differences in the length of the vocal tract. These differences have been leveraged in gender classification through speech analysis based on pitch and formant frequencies [36], HNR [37], and linear predictive components and Mel Frequency Cepstral Coefficients (MFCC) [38]. Previous work found differences in speech depending on gender in the effects of depression and the effectiveness of classifiers for its detection [39]. This is why this study considers males and females separately.

Statistical analysis

The data collected was described using mean and standard deviation for quantitative variables, and frequency and percentage for qualitative variables. Demographic characteristics such as age and gender were compared between different groups of caregivers using a Chi-squared test for qualitative variables (e.g., gender) and an analysis of variance (ANOVA) performed for quantitative variables (e.g., age). Similarly, the data measured for voice and scores were compared between different groups of caregivers. The normality of the collected data was tested using a Shapiro test. In order to test the relationship between the different voice measures and the measured scores Spearman correlations was used. In addition, to test the link between the voice measures and the therapist's decision, Student t-tests or Wilcoxon-Mann-Whitney tests were performed. A p-value of <0.05 was considered significant. The analyses were performed using the free statistical software R Studio v4.0.0 [40]. Further, regression analyses were performed with the extracted vocal features to determine the error rate for predicting the participants’ stress score.

Results

Participants

In total, 89 French speaking health professionals aged between 20-74 accepted the phone call and their speech samples were recorded and analyzed. Their demographic characteristics are presented in Table 1.

Table 1. Descriptive statistics for participants (N=89)

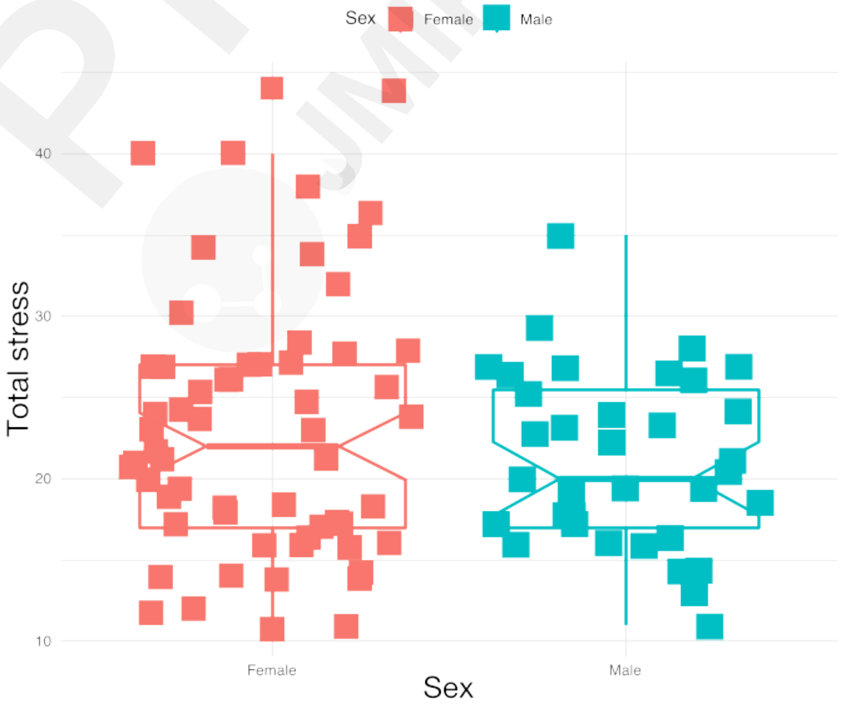
	n	%	Male		Female		p-value*
			n	%	n	%	
Gender							
M	31	34.8					
F	58	65.2					
Education							0.027
< 12 years	19	23.5	1	3.6	18	34.0	
> 12 years	62	76.5	27	96.4	35	66.0	
called during or after lockdown							0.027
During	34	38.2	7	22.6	27	46.6	
After	55	61.8	24	77.4	31	53.4	
Perceived Stress Questionnaire							0.471
Knows how to manage stress (<21)	40	44.9	16	51.6	24	41.4	
Generally knows how to cope with stress (21-26)	25	28.1	9	29.0	16	27.6	
Life is a constant threat (>26)	24	27.0	6	19.4	18	31.0	
Motivation Stress Affect scale							0.995
Score=0	23	25.8	8	25.8	15	25.9	
Score >0	66	74.2	23	74.2	43	74.1	
MSA_motivation							0.466
Score=0	30	33.7	12	38.7	18	31.0	
Score >0	59	66.3	19	61.3	40	69.0	
MSA_depression							0.320
Score=0	57	60	22	71.0	35	60.3	
Score >0	32	36	9	29.0	23	39.7	
Follow up request							0.362

No	74	84.1	28	90.3	46	80.7
Yes	14	15.9	3	9.7	11	19.3

*Chi-square test or exact Fisher test

The mean age was 40.53 (Standard Deviation (SD)=14.19); the mean stress score on the PSQ was 22.43 (SD=7.16) and on the MSA questionnaire 2.92 (SD=2.09). The majority of participants scored below 26 on the PSQ but above 0 on the MSA. Results on the PSQ and MSA Stress are proportional. We found that 27% (24 out of 89 participants) of the recorded health professionals experience intense stress, and 28.1% occasional stress (25 out of 89 participants). Only 15.9% (14 out of 89 participants) requested a follow-up. The stress level is gender dependent with females reporting higher stress levels. For males, stress level tends to drop with age. Figure 1. shows a distribution of the total stress score across genders. The total stress score in the female group is more dispersed than in the male counterpart and is generally higher. 14 participants (11 females, 3 males) asked for a follow up call. Their mean PSQ (31.78) and MSA (5.57) scores were significantly higher than those who did not ask for a follow up with a mean PSQ score of 20.60 and MSA of 2.38.

Figure 1. Stress score distribution



Correlations

First, vocal and non-vocal features were analyzed in relation to the stress level. The data set it quite small and therefore, rather than training a classifier, we performed correlation analysis between the features computed for each speech task and the reported stress level. Further, only extracted speech features were considered (a-priori, non-meaningful and features, like the id were removed).

We performed a selection of top-k features based on their descriptive power for the target variable ‘total stress score’. Vocal features might be gender-dependent. Therefore, we performed a selection of top features for male and female datasets separately. We used Spearman correlation, since we had both ordinal and continuous features (the target ‘Stress total score’ is ordinal). Since Spearman correlation uses only the ranks of the variables and not their raw values, we could omit the normalization step. We considered absolute values of the correlation coefficient for feature scoring. Results are presented in Table 2.

Table 2. Correlation between stress levels and speech features

Top-10 features for female dataset	Task	Spearman correlation
MFCC3 acceleration skewness	Pos	0.49
MFCC2 mean	Neu	0.44
Pitch range	Pos	0.44
MFCC3 acceleration skewness	Neg	0.43
MFCC2 mean	Pos	0.44
MFCC5 acceleration kurtosis	Neg	-0.42
MFCC2 mean	Neg	0.43
MFCC5 velocity kurtosis	Neg	-0.40
MFCC3 acceleration skewness	Neu	0.39
MFCC5 velocity kurtosis	Neg	0.39
Top-10 features for male dataset	**Task**	**Spearman correlation**
Upper quartile F0	Neu	-0.54
Pronunciation posteriori probability score percentage	Pos	-0.50

Energy acceleration mean	Pos	0.52
Mean F0	Neu	-0.51
MFCC9 kurtosis	Pos	0.41
MFCC9 variance	Pos	-0.44
Upper quartile F0	Neg	-0.47
MFCC4 acceleration mean	Pos	-0.40
Upper quartile F0	Pos	-0.47
MFCC12 acceleration skewness	Neu	-0.42

Pos: positive story; neg: negative story; neu: neutral story

The main speech parameters correlating with stress levels in both genders were spectral (formant) features and namely the Mel-frequency cepstral coefficients (MFCC). These features are characterizing the spectrum of speech, which is the frequency distribution of the speech signal at a specific time. MFCCs derived by computing a spectrum of the log-magnitude Mel-spectrum of the audio segment. The lower coefficients represent the vocal tract filter and the higher coefficients represent periodic vocal fold sources [18]. Moreover, in males' prosodic characteristics such as the fundamental frequency and in females in the positive storytelling, pitch ranges were associated with stress levels.

For female participants, correlation analysis between negative, positive, and neutral features and the target feature 'Stress total score' was performed. In the top-5 features, we have MFCC acceleration skewness, which correlates with the stress level by 0.45 and 0.37 in the positive and neutral tasks. The other features in top-5 are task-specific. Thus, for each task there are a different set of features associated with stress level.

For male participants, the selection was performed analogously. The top features are task-specific as well and they differ from the features for the female dataset. In this sample, we obtained more negatively correlating features than for the female data meaning that features for instance related to F0 of low value (Mean F0 in the neutral story with -0.51, Upper quartile F0 in the negative and positive story with -0.47 and in the neutral story with -0.54) are associated with high stress scores. Low values represent in general a smaller pitch range.

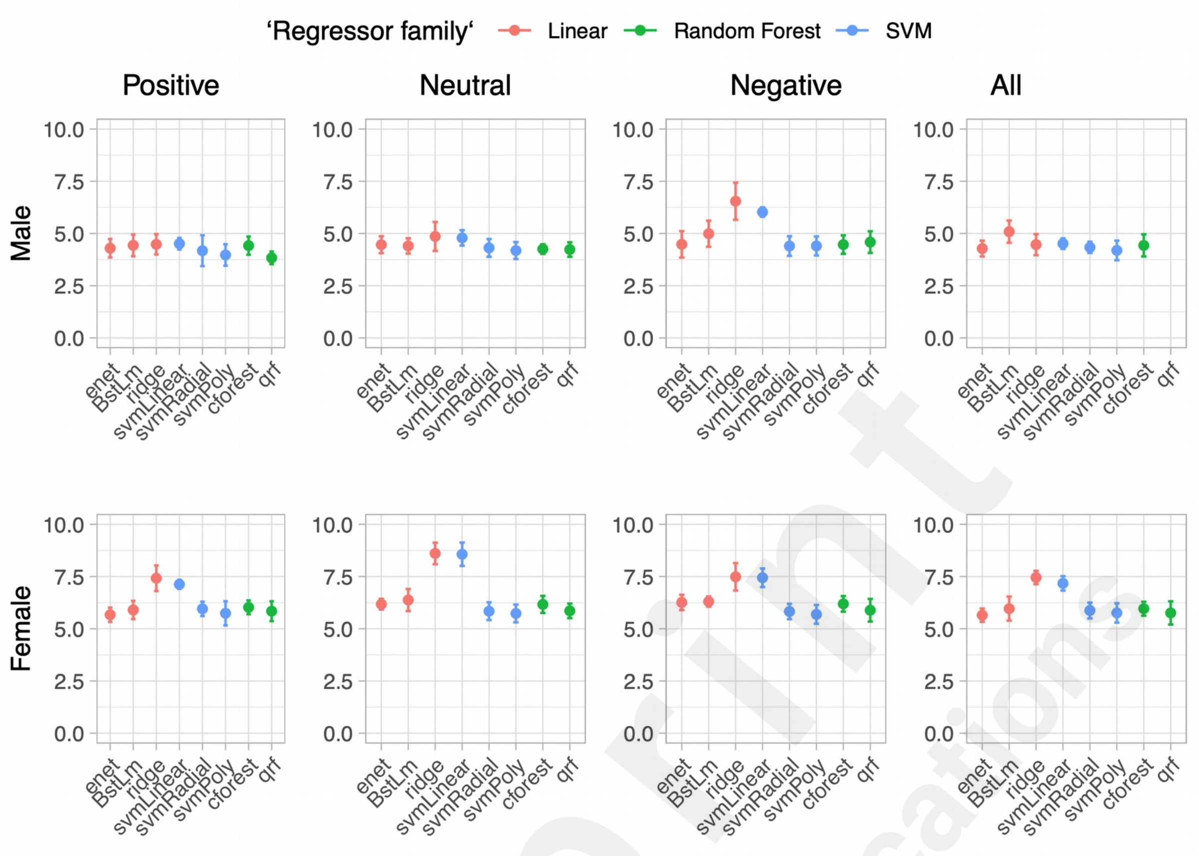
Regression

Stress scores were regressed against measurements for positive, neutral and negative tasks. Similarly, the regression for tasks of different sentiments was performed for groups of female and male participants to allow for possible impacts of genders on stress levels. For the regressors, we used linear, support vector machine (SVM) and random forest regressors to predict the stress scores.

The first approximates the stress score by estimating coefficients for each feature in the training data, where greater coefficients indicate a greater influence over the predicted value. Linear regression models are fast, highly interpretable and commonly used for prediction of stress scores from audio features and speech analysis according to previous studies [41-43]. The random forest regressor creates a number of decision trees that are constructed based on random sampling from the training data; each tree then attempts to determine the best way to predict the scores given the data it receives. Each decision tree outputs a predicted value and the mode value is selected. Decision trees methods have shown high accuracy with good interpretability in similar studies where vocal and linguistic features were employed for detection of emotions, social signals and mental health problems [44-46]. The SVM regressor takes each set of features and projects them as a vector onto a space and attempts to find the optimal way to separate the data. The stress score is then based on the distance from that separator. Stress modelling with inputs from physiological sensors or audio sources using SVM has also been reported to give high model performance previously [47-49]. In recent studies, both SVM and random forest provided notably high prediction/classification strength for stress detection using various speech features [50-52].

The R package caret is used for data training and validation. A 10-fold cross validation is performed and performance is evaluated using the mean absolute error (MAE); the average of the absolute difference between the predicted and actual value from our models for all participants. The score ranges from 0 to infinity, where a score closer to 0 indicates a better fitting model.

Figure 2. Performances of different computerized regression models in predicting stress levels based on vocal features



The prediction of total stress score using all or a subset of tasks in male or female subjects were carried out using various baseline regression models, whose performance was evaluated by the plot (Figure 2.) where the mean absolute errors are presented by the y-axis. Overall, the prediction strength in males is better than in females for all sentiments as shown by a trend of lower errors (male: lowest MAE with 3.84; females: lowest MAE with 5.56). It is notable that stress score regression models based on negative tasks in males and neutral tasks in females performed relatively poorly compared to other tasks. For both male and female participants, using positive tasks for regression yielded equivalent or better results than using all tasks, suggesting a subset of tasks could be employed for accurate and less time-consuming prediction of stress score. An overview of the lowest scores for each testing scenario is presented in Table 3.

Table 3. Mean Absolute Error - the Lowest Scores for Each testing Scenario

	Positive	Neutral	Negative
All	5.31(0.25) enet - Linear	**5.25(0.28)** qrf - Random Forest	5.34(0.35) svmPoly - SVM
Male	**3.84(0.43)** qrf - Random Forest	4.40(0.37) BstLm - Linear	4.37 (0.43) svmPoly - SVM

Female	**5.56 (0.41)** enet - Linear	5.84(0.42) svmRadial - SVM	5.68 (0.45) svmPoly - SVM

All regression models outperformed their respective baseline MAEs (4.46 and 6.35 in male and female respectively). Linear models and SVM are most precise for the prediction of total stress score in general.

Discussion

Principle findings

The purpose of this study was to investigate the potential of using automatic speech analysis for the detection of stress in healthcare professionals during the current COVID-19 pandemic. This would potentially lead to earlier and timely prevention among this at high-risk population. For this, firstly, speech samples were collected over the phone and various voice features extracted and compared with classical stress measures. Secondly, based on the extracted features participants' obtained scores on the completed stress scale were predicted.

The main outcomes of this study are, that we demonstrated the feasibility under the given context, as all participants were collaborative and appreciated the initiative of rapidly applying the use of this existing technology to this specific use case. Moreover, from phone call recordings, a number of vocal correlates of stress have been identified, namely in the area of spectral features (MFCC) as well as prosodic features such as fundamental frequency (F0) which seem the most commonly reported in well-controlled trials [11]. Stress scores could be predicted based on speech features with relatively small errors.

Spectral features characterize the speech spectrum; the frequency distribution of the speech signal at a specific time instance information in some high dimensional representation [18]. They capture information regarding changes in muscle tension and control and have consistently been observed to change with a speaker's mental state. A few depression studies reported a relative shift in energy with increasing depression severity [53-54].

Another result we obtained was that most identified vocal features are task- as well gender-dependent. Interestingly, in the female group, MFCC features seemed to be associated with stress levels during all tasks, meaning that it did not matter what they were talking about as long as sufficient speech is captured, meaningful information could be extracted and subtle signs of stress level detected. In return, in the male dataset, Upper F0 appeared as a task independent feature sensitive to stress levels. Overall in this set, we observed more features with negative correlation rather than for the female data.

Voice production can be divided into three processes: breathing, phonation and resonance stress [55]. For the second process, phonation, the vocal folds must close and open again to create vibration. The frequency rate of these pulses determines the fundamental frequency (F0) of the vocal source contributing to the perceived pitch of the sound.

Previous research showed that increased muscle tension tends to be caused by stress [56-57] resulting in a tensing of the vocal folds which in turn most likely causes a raising of fundamental frequency (F0). A recent review on voice analysis in stress [22] states that the parameter F0 has been considered as a ‘universal stress indicator’ whereas increased levels of F0 might be linked with acute bottom up processes of sympathetic arousal. Similar studies of analysis of phone call recordings during situational stress situations revealed an increase in F0 and intensity with presumed levels of stress [58-59;56]. Our findings seem consistent with the majority of acoustic studies, pointing to F0 as one important marker of stress levels.

However, most correlation we found were with resonance (formant) parameters which are involved in the quality of sound shaping and vowel and consonant pronunciation and produced by the muscle activity involved in the shaping of the resonant cavities of the vocal tract system. [60]. These parameters are less documented in regards to stress. MFCC in particular can be indicative of breathiness in the voice [61]. Interestingly, one study found a circadian pattern in MFCCs due to sleep deprivation. For this, voice perturbations were compared with classical sleep measures [62] and correlation found between fatigue score and MFCCs. This might eventually explains our results as most participants reported as well signs of fatigue during the interviews.

Another study examined speech in students under exam stress and a few days later for which heart rate was measured to control for the actual stress levels. Under stress, heart rate increased, F0 and F0 SD increased, F1 and F2 increased and MFCCs decreased in relation to a baseline [63].

It can be hypothesized that given our recorded population reported relatively mild to moderate levels of stress, rather subtle changes in voice parameters were found and therefore weaker correlations. However, it is important to underline that changes in features we found to be sensitive to stress levels are gender but not necessarily task-dependent, and most likely too small to be detectable by the human ear; but captured by the automatic speech analysis. We assume that by applying this technology to regular check-up calls of people exposed to high stress levels such as health care professionals, very early signs can be detected in their voice allowing for timely preventive strategies.

Regression models using vocal features performed relatively well in predicting stress scores namely in the positive story task for both genders (average of 5.3 points of errors). It shows that the technology would capture indicative patterns from even a short amount of time, possible even from one task,, to recognize tendencies of stress levels in a fragile but healthy population; representing a promising rapid tool for prediction of stress scores.

Strengths of this study

This study is a first step into early identification of at-risk population, such as caregivers, who do not directly express their psychological suffering. We can imagine extending this technique to other fragile populations for early screening such as teenagers who are victims of school harassment, or women who are victims of abuse, where timely management could potentially prevent the development of co-morbidities such as depression and anxiety. Moreover, patient populations who have difficulty expressing their problems such as autism spectrum disorder, or dementia patients, could benefit from this technology.

Generally, remote psychological counseling is controversial. Nevertheless, it is becoming necessary due to current economic, social and health constraints, but has been received by professionals and patients with mixed feelings. Indeed, the non-verbal part of the communication is lost and the dynamics of interaction are not the same. However, contrary to these preconceived ideas, we have noticed during this work that it is easier for certain participants to open up and speak about personal issues during these interviews in a liberating manner similar to a confessional. Not being in the physical presence of the listener may facilitate persona expression with less fear of being judged. This aspect is very interesting during a screening because it considerably accelerates the process of detection and diagnosing psychological symptoms.

Weaknesses of the study

This project has been rapidly implemented initially with an approach of qualitative and quantitative data analysis that should contribute to early and timely assistance of health professionals during the COVID-19 pandemic. The staff available to participate in the study was limited. Patient selection has been done on a voluntary basis. It is conceivable that the population studied was more concerned about their state of psychological suffering and therefore potentially has a selection bias.

Although the voice recordings were made in the middle of the interview without this time being precisely stated, it is possible that some patients may have suspected this, which could have been anxiety-provoking and skewed our results. Recording throughout the interview for parameters not affected by the tasks would provide more data and more robust results.

Finally, the obtained correlations can be considered as rather moderate which makes it difficult to draw any strong conclusions. A larger dataset, ideally of longitudinal nature, with more precise characterization of the speakers is needed in order to verify if the correlating features represent real markers of stress.’

Future perspective

For future work, we propose to perform this analysis on a larger data set and to build a prediction model. In case of insufficient number of observations per stress level, the number of stress levels can be reduced by binning. Binning can also be carried out on characteristic values.

Further studies with acoustic measurements and stress questionnaires at regular time intervals would allow analysis of the kinetics of the markers and a better perception of their sensitivity and specificity. In addition, adding clinical measurements of psychiatric symptoms (such as the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition -DSM-5) [64] would make it possible to perceive whether one of the markers is predictive of an anxiety or depression disorder. The use of the tool could be combined with the delivering of preventive strategies such as physical exercises, adaptation of diet, psychotherapy, meditation or the use of symptomatic treatments and employed at the same for the evaluation of the obtained effects. However, in order to produce a real world application of this technology, larger validation studies have to be performed to demonstrate clinical meaningfulness by comparing its performance to standardized measurement tools.

Authors' Contributions

AK, KR, JE and PR designed and conducted the study. NL, HL and RF analyzed the data. KR and AK, NL, HL and PR drafted the manuscript. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

This research was supported by a grant from the fondation JL Noisiez and the association IA. Thanks to all professional healthcare workers who accepted to participate in this study.

Conflicts of Interest

None declared.

References:

1. Wuhan Municipal Health Committee. Report of Wuhan Municipal Health Committee on viral pneumonia of unknown cause [in Chinese] URL: http://wjw.wuhan.gov.cn/front/web/showDetail/2020011109035
2. World Health Organization. 2020 Jan 31. Novel Coronavirus (2019-nCoV) Situation Report - 11 URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200131-sitrep-11-ncov.pdf?sfvrsn=de7c0f7_4
3. Liu X, Kakade M, Fuller CJ, et al. Depression after exposure to stressful events: Lessons learned from the severe acute respiratory syndrome epidemic. Compr Psychiatry 2012; 53: 15–23
4. Lung FW, Lu YC, Chang YY, Shu BC. Mental symptoms in different health professionals during the SARS attack: A Follow-up study. Psychiatr Q 2009; 80: 107–16. 5
5. Wu P, Fang Y, Guan Z, et al. The Psychological Impact of the SARS Epidemic on Hospital Employees in China: Exposure, Risk Perception, and Altruistic Acceptance of Risk.
6. Pappa S, Ntella V, Giannakas T, Giannakoulis VG, Papoutsi E, Katsaounou P. Prevalence of depression, anxiety, and insomnia among healthcare workers during the COVID-19 pandemic: A systematic review and meta-analysis. Brain Behav Immun. 2020;88:901-907. doi:10.1016/j.bbi.2020.05.026
7. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav. 1983 Dec;24(4):385-96. PMID: 6668417.

8. Roohafza H, Ramezani M, Sadeghi M, Shahnam M, Zolfagari B, Sarafzadegan N. Development and validation of the stressful life event questionnaire. Int J Public Health. 2011 Aug;56(4):441-8.
9. Amirkhan JH. Stress overload: a new approach to the assessment of stress. Am J Community Psychol. 2012 Mar;49(1-2):55-71. doi: 10.1007/s10464-011-9438-x. Erratum in: Am J Community Psychol. 2012 Mar;49(1-2):72.
10. Petrowski K, Paul S, Albani C, Brähler E. Factor structure and psychometric properties of the trier inventory for chronic stress (TICS) in a representative German sample. BMC Med Res Methodol. 2012 Apr 1;12:42.
11. Giddens, C. L., K. W. Barron, J. Byrd-Craven, K. F. Clark and A. S. Winter (2013). Vocal indices of stress: a review. J Voice 27(3): 390 e321-399.
12. Pisanski, K., J. Nowak and P. Sorokowski (2016). Individual differences in cortisol stress response predict increases in voice pitch during exam stress. Physiol Behav 163: 234-238.
13. Kirchhübel, C., D. M. Howard and A. W. Stedmon (2011). Acoustic correlates of speech when under stress: Research, methods and future directions. The International Journal of Speech, Language and the Law 18(1): 75-98.
14. Hollien, H. (2014). Vocal fold dynamics for frequency change. J Voice 28(4): 395-405.
15. Sobin, C and Sackheim, H. A. Psychomotor symptoms of depression. American Journal of Psychiatry, vol. 154, no. 1, pp. 4–17, 1997.
16. Schrijvers, D, Hulstijn, W and Sabbe, BG. Psychomotor symptoms in depression: a diagnostic, pathophysiological and therapeutic tool. Journal of Affective Disorders, vol. 109, no. 1-2, pp. 1–20, 2008.
17. Bylsam, LM, Morris, BH and Rottenberg, J . A meta-analysis of emotional reactivity in major depressive disorder. Clinical Psychology Review, vol. 28, pp. 676–691, 2008.
18. Cummins, N, Scherer, S, Krajewski, J, Schnieder, S, Epps, J and Quatieri, T. A review of depression and suicide risk assessment using speech analysis. Speech Communication, vol. 17, pp. 10–49, 2015.
19. Nilsonne, A. Speech characteristics as indicators of depressive illness. Acta Psychiatrica Scandinavica, vol. 77, no. 3, pp. 253–263, 1988.
20. Leff, J and Abberton, E. Voice pitch measurements in schizophrenia and depression. Psychological Medicine, vol. 11, no. 4, pp. 849–852, 1981.
21. Marmar CR, Brown AD, Qian M, et al. Speech-based markers for posttraumatic stress disorder in US veterans. Depress Anxiety. 2019;36(7):607-616. doi:10.1002/da.22890

22. Van Puyvelde M, Neyt X, McGlone F and Pattyn N (2018) Voice Stress Analysis: A New Framework for Voice and Effort in Human Performance. Front. Psychol. 9:1994. doi: 10.3389/fpsyg.2018.0199
23. Robert P, Lanctôt KL, Agüera-Ortiz L, et al. Is it time to revise the diagnostic criteria for apathy in brain disorders? The 2018 international consensus group. Eur Psychiatry. 2018;54:71-76. doi:10.1016/j.eurpsy.2018.07.008
24. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res. 1983;17:37–49.
25. König A, Linz N, Zeghari R, Klinge X, Tröger J, Alexandersson J, Robert P. Detecting Apathy in Older Adults with Cognitive Disorders Using Automatic Speech Analysis. J Alzheimers Dis. 2019;69(4):1183-1193.
26. Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A Global Measure of Perceived Stress. Journal of Health and Social Behavior, 24(4), 385-396.
27. URL: https://ki-elements.de
28. Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.
29. URL: https://github.com/Shahabks/my-voice-analysis
30. Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International* **5:9/10**, 341-345.
31. URL: https://github.com/jameslyons/python_speech_features
32. URL: https://pypi.org/project/librosa/ version 0.7.2
33. Kreiman, J., & Gerratt, B. R. (2005). Perception of aperiodicity in pathological voice. The Journal of the Acoustical Society of America, 117(4), 2201-2211.
34. Michaelis, D., Fröhlich, M., Strube, H. W., Kruse, E., Story, B., & Titze, I. R. (1998, June). Some simulations concerning jitter and shimmer measurement. In 3rd International Workshop on Advances in Quantitative Laryngoscopy, Aachen, Germany (pp. 744-754).
35. Murphy, P. J., & Akande, O. O. (2005, April). Cepstrum-based estimation of the harmonics-to-noise ratio for synthesized and human voice signals. In International conference on nonlinear analyses and algorithms for speech processing (pp. 150-160). Springer, Berlin, Heidelberg.
36. Childers DG, Wu K (1991) Gender recognition from speech. Part II: Fine analysis. J Acoust Soc Am 90, 1841-1856.

37. Heffernan K (2004) Evidence from HNR that/s/is a social marker of gender. Toronto Working Papers in Linguistics, 23.
38. Wu K, Childers DG (1991) Gender recognition from speech. Part I: Coarse analysis. J Acoust Soc Am 90, 1828-1840.
39. Low, LSA, Maddage NC, Lech M, Sheeber LB, Allen NB (2011) Detection of clinical depression in adolescents' speech during family interactions. IEEE Trans Biomed Eng 58, 574–586.
40. URL: https://cran.r-project.org/
41. Gillespie, S., Moore, E., Laures-Gore, J., Farina, M., Russell, S., & Logan, Y. Y. (2017, March). Detecting stress and depression in adults with aphasia through speech analysis. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5140-5144). IEEE.
42. van den Broek, E. L., van der Sluis, F., & Dijkstra, T. (2010). Telling the story and re-living the past: How speech analysis can reveal emotions in post-traumatic stress disorder (PTSD) patients. In Sensing emotions (pp. 153-180). Springer, Dordrecht.
43. Muaremi, A., Arnrich, B., & Tröster, G. (2013). Towards measuring stress with smartphones and wearable devices during workday and sleep. BioNanoScience, 3(2), 172-183.
44. Hasan, M., Rundensteiner, E., & Agu, E. (2014). Emotex: Detecting emotions in twitter messages.
45. Gosztolya, G., Busa-Fekete, R., & Tóth, L. (2013). Detecting autism, emotions and social signals using AdaBoost. Interspeech.
46. McCabe, R., Howes, C., & Purver, M. (2014, June). Linguistic indicators of severity and profess in online text-based therapy for depression. Association for Computational Linguistics.
47. Al-Shargie, F., Tang, T. B., Badruddin, N., & Kiguchi, M. (2018). Towards multilevel mental stress assessment using SVM with ECOC: an EEG approach. Medical & biological engineering & computing, 56(1), 125-136.
48. Rabaoui, A., Davy, M., Rossignol, S., & Ellouze, N. (2008). Using one-class SVMs and wavelets for audio surveillance. IEEE Transactions on information forensics and security, 3(4), 763-775.
49. Chang, C. Y., Chang, C. W., Zheng, J. Y., & Chung, P. C. (2013). Physiological emotion analysis using support vector regression. Neurocomputing, 122, 79-87.
50. Soury, M., & Devillers, L. (2013, September). Stress detection from audio on multiple

window analysis size in a public speaking task. In 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (pp. 529-533). IEEE.

51. Sysoev, M., Kos, A., & Pogačnik, M. (2015). Noninvasive stress recognition considering the current activity. Personal and Ubiquitous Computing, 19(7), 1045-1052.
52. Gjoreski, M., Gjoreski, H., Lutrek, M., & Gams, M. (2015, July). Automatic detection of perceived stress in campus students using smartphones. In 2015 International Conference on Intelligent Environments (pp. 132-135). IEEE.
53. Cummins, N., Epps, J., Ambikairajah, E., 2013a. Spectro-temporal analysis of speech affected by depression and psychomotor retardation. Proceedings of ICASSP. IEEE, Vancouver, Canada, pp. 7542–7546.
54. Cummins, N., Epps, J., Sethu, V., Breakspear, M., Goecke, R., 2013b. Modeling spectral variability for the classification of depressed speech. Proceedings of Interspeech. ISCA, Lyon, France, pp. 857–861.
55. Kreiman, J., and Sidtis, D. (2011). Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception. Hoboken, NJ: JohnWiley & Sons. doi: 10.1002/9781444395068
56. Streeter, L. A., N. H. Macdonald, W. Apple, R. M. Krauss and K. M. Galotti (1983). Acoustic and perceptual indicators of emotional stress. J Acoust Soc Am 73(4): 1354-1360.
57. Scherer, K.R., Grandjean, D., Johnstone, T., Klasmeyer, G. and Bänziger, T. (2002) Acoustic correlates of task load and stress. Proceedings, ICSLP 2017–2020. Denver, USA.
58. Ruiz, R., Absil, E., Harmegnies, B., Legros, C., and Poch, D. (1996). Time and spectrum-related variabilities in stressed speech under laboratory and real conditions. Speech Commun. 20, 111–129. doi: 10.1016/S0167-6393(96)
59. Jessen, M. (2006) Einfluss von Stress auf Sprache und Stimme. Unter besondere Berücksichtigung polizeidienstlicher Anforderungen. Idstein: Schulz-Kirchner Verlag
60. Gopalan, K., Wenndt, S., and Cupples, E. J. (1999). An analysis of speech under stress using certain modulation features. IECON'99, in Proceedings of the 25th Annual Conference of the IEEE Industrial Electronics Society, San Jose, CA. doi: 10.1109/iecon.1999.819381
61. Hillenbrand, J., and Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. J. Speech Lang. Hear. Res. 39311–321. doi: 10.1044/jshr.3902.311
62. Greeley, H. P., Friets, E., Wilson, J. P., Raghavan, S., Picone, J., and Berg, J. (2006). Detecting fatigue from voice using speech recognition, in Proceedings of the IEEE

International Symposium on Signal Processing and InformationvTechnology, (Louisville, KY: IEEE), 567–571. doi: 10.1109/ISSPIT.2006.27

63. Sigmund, M. (2006). Introducing the database ExamStress for speech under stress, in Proceedings of the 7th Nordic Signal Processing Symposium, 2006.NORSIG 2006, (Reykjavik: IEEE), 290–293. doi: 10.1109/NORSIG.2006.275258
64. American Psychiatric Association (2013). Diagnostic and Statistical Manual of Mental Disorders (Fifth ed.). Arlington, VA: American Psychiatric Publishing. pp. 5–25. ISBN 978-0-89042-555-8.

Supplementary Files

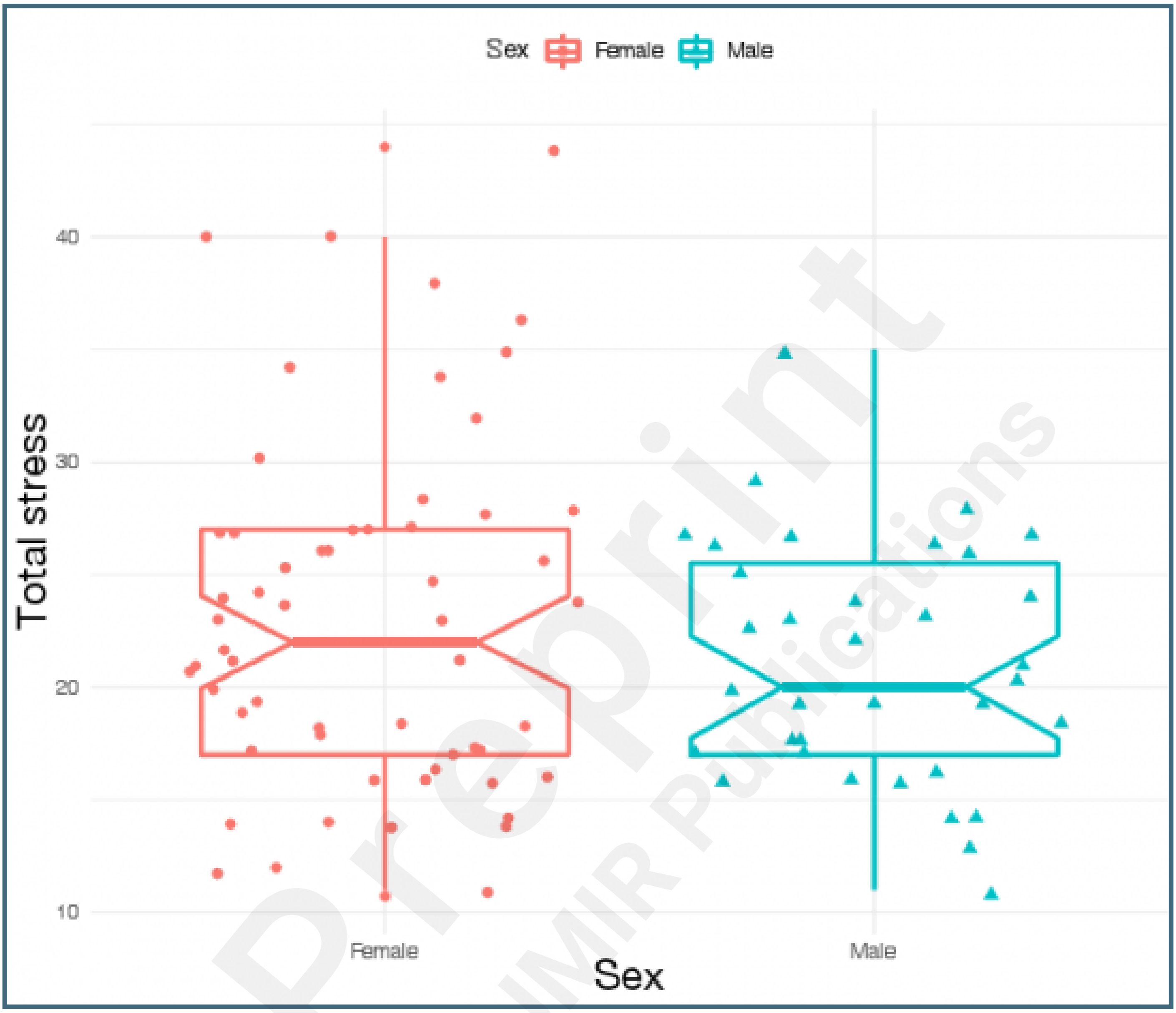
Revised manuscript with track change.
URL: http://asset.jmir.pub/assets/fb9ea5f4379a639305889323c3299568.docx

Responses to reviewers.
URL: http://asset.jmir.pub/assets/b4277b8abc6d5c9dbedf704c41416d72.docx

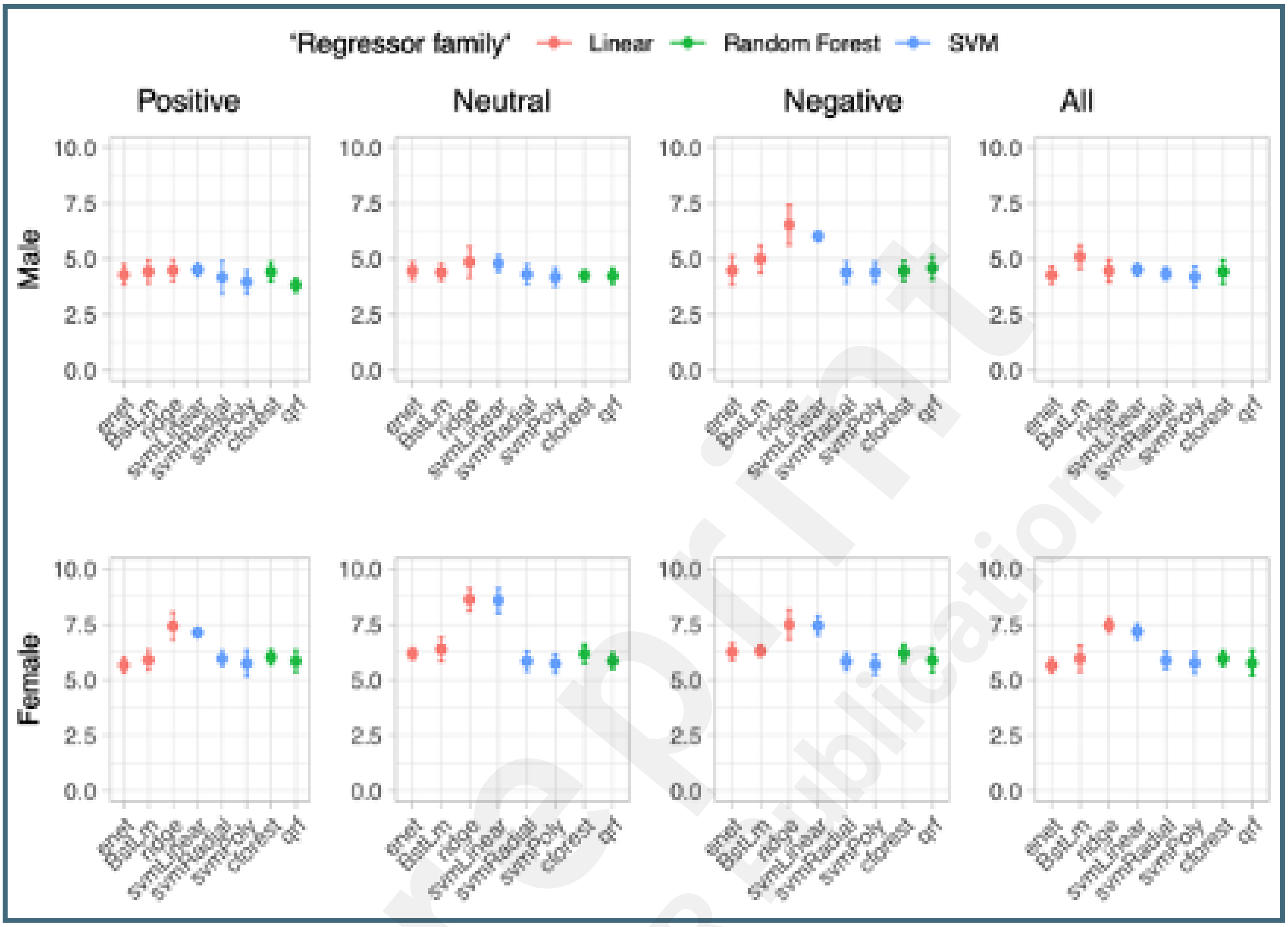
Untitled.
URL: http://asset.jmir.pub/assets/bb33dbc94d55e9b7987527027c61c98a.docx

Figures

Stress score distribution across genders.



Performances of different computerized regression models in predicting stress levels based on vocal features.


'Regressor family'
Linear
Random Forest
SVM
Positive
Neutral
Negative
All
Male
Female
enet
BstLm
ridge
svmLinear
svmRadial
svmPoly
cforest
qrf