Protocol

Associations Between the Severity of Obsessive-Compulsive Disorder and Vocal Features in Children and Adolescents: Protocol for a Statistical and Machine Learning Analysis

Line Katrine Harder Clemmensen[1*], PhD; Nicole Nadine Lønfeldt[2*], PhD; Sneha Das[1], PhD; Nicklas Leander Lund[3], MSc; Valdemar Funch Uhre[2,4], PhD; Anna-Rosa Cecilie Mora-Jensen[2,5], MD; Linea Pretzmann[2,5], MSc; Camilla Funch Uhre[5,6], PhD; Melanie Ritter[2], MSc; Nicoline Løcke Jepsen Korsbjerg[2], MSc; Julie Hagstrøm[2], PhD; Christine Lykke Thoustrup[2,5], MSc; Iben Thiemer Clemmesen[2], MSc; Kersten Jessica Plessen[7], MD, PhD; Anne Katrine Pagsberg[2,5], MD, PhD

[1]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Copenhagen, Denmark

[2]Child and Adolescent Mental Health Center, Copenhagen University Hospital, Mental Health Services Copenhagen, Copenhagen, Denmark

[3]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby, Denmark

[4]Centre for Functional and Diagnostic Imaging and Research, Danish Research Centre for Magnetic Resonance, Copenhagen University Hospital, Amager and Hvidovre, Denmark

[5]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[6]Center for Clinical Neuropsychology, Children and Adolescents, Rigshospitalet, Copenhagen, Denmark

[7]Division of Child and Adolescent Psychiatry, Department of Psychiatry, Lausanne University Hospital, Centre Hospitalier Universitaire Vaudois, University of Lausanne, Lausanne, Switzerland

[*]these authors contributed equally

Corresponding Author:
Line Katrine Harder Clemmensen, PhD
Department of Applied Mathematics and Computer Science
Technical University of Denmark
Richard Petersens Plads
Bygning 324
Copenhagen, 2800
Denmark
Phone: 45 45 25 37 64
Email: lkhc@dtu.dk

Abstract

Background: Artificial intelligence tools have the potential to objectively identify youth in need of mental health care. Speech signals have shown promise as a source for predicting various psychiatric conditions and transdiagnostic symptoms.

Objective: We designed a study testing the association between obsessive-compulsive disorder (OCD) diagnosis and symptom severity on vocal features in children and adolescents. Here, we present an analysis plan and statistical report for the study to document our a priori hypotheses and increase the robustness of the findings of our planned study.

Methods: Audio recordings of clinical interviews of 47 children and adolescents with OCD and 17 children and adolescents without a psychiatric diagnosis will be analyzed. Youths were between 8 and 17 years old. We will test the effect of OCD diagnosis on computationally derived scores of vocal activation using ANOVA. To test the effect of OCD severity classifications on the same computationally derived vocal scores, we will perform a logistic regression. Finally, we will attempt to create an improved indicator of OCD severity by refining the model with more relevant labels. Models will be adjusted for age and gender. Model validation strategies are outlined.

Results: Simulated results are presented. The actual results using real data will be presented in future publications.

Conclusions: A major strength of this study is that we will include age and gender in our models to increase classification accuracy. A major challenge is the suboptimal quality of the audio recordings, which are representative of in-the-wild data and a large body of recordings collected during other clinical trials. This preregistered analysis plan and statistical report will increase the validity of the interpretations of the upcoming results.

International Registered Report Identifier (IRRID): DERR1-10.2196/39613

(JMIR Res Protoc 2022;11(10):e39613) doi: 10.2196/39613

KEYWORDS

machine learning; obsessive-compulsive disorder; vocal features; speech signals; children; teens; adolescents; OCD; AI; artificial intelligence; tool; mental health; care; speech; data; clinical trial; validity; results

Introduction

Obsessive-compulsive disorder (OCD) is a chronic, debilitating disorder, which can lower self-esteem, shorten life-expectancy, strain the family, and make it difficult to maintain friendships and attend school [1,2]. First-line treatment for moderate to severe OCD in youth (defined as individuals under age 18 years of age) is cognitive behavioral therapy (CBT) with exposure and response prevention (ERP) [3,4]. During exposure practice, the child tracks the level of distress caused by symptom-provoking stimuli across and within sessions. Here, distress refers to fear, disgust, discomfort, shame, embarrassment, and feelings of incompleteness or emptiness. Monitoring distress provides useful information to clinicians, who use it to plan exposures and help the patient remain mentally present with the exposure [5]. Distress levels also comprise one dimension of the gold-standard measure of symptom severity in OCD. When collected over time, distress can provide information about disease progression and improvement [6]. Frequent measures of distress are essential for understanding mechanisms of change in exposure-based therapies [5]. Self-rated distress may not be frequent enough to discover the processes responsible for therapeutic change, which has led some researchers to code videos of exposure sessions [7]. Behavioral coding is a time-consuming, costly process that is prone to inconsistency and not entirely immune to bias. Ideally, distress would be assessed objectively and affordably in a noninvasive manner.

Objective and automatic psychiatric assessments can be achieved by feeding vocal features into machine learning models. Speech patterns, tempo, volume, and intonation comprise an important part of the overall clinical impression that has been used to diagnose psychiatric disorders for at least 100 years [8]. Speech reflects changes in cognition, affect, motor characteristics, and physiology seen in psychiatric disorders [9]. Voice quality features capture information relating to voice creakiness, harshness, and breathiness [9]. Decreased formant frequencies observed in depressed and anxious speech may reflect dry mouth, decreased articulation, or motor coordination [9].

Vocal features have demonstrated promising results in machine learning for predicting the severity of psychiatric disorders and clinical improvement. A recent review summarized 127 studies that have used automatically extracted speech features to detect the presence or severity of psychiatric disorders [10]. The vocal fold features, jitter, and shimmer were found to be significantly elevated in adults with depression, anxiety, and OCD. Only one study has investigated vocal fold features in OCD. Results from 35 adults suggest that individuals with OCD have voices with more jitter, breathiness, hoarseness and speak at a lower rate than individuals without a psychiatric diagnosis [11]. Increased percent jitter and hoarseness have also been found in children (6-15 years old) with attention-deficit/hyperactivity disorder compared with healthy controls [12].

One major conclusion of the systematic review is that future studies should focus on linking vocal features to specific transdiagnostic problems, such as distress [10]. One study found that self-reported distress and electrodermal activity corresponded with vocal indicators of distress [13]. High levels of cortisol have also been linked to vocal indicators of distress [14]. Another study used audio recordings of 3- to 8-year-old children giving a speech under stressful conditions. The machine learning model classified children with and those without internalizing disorders using vocal features as input with 80% accuracy [15].

The aim of this paper is to increase the robustness of a planned study by documenting our analysis plan of a study with the following objectives: (1) to investigate the correspondence between activation-specific features and OCD diagnosis and severity [16] and (2) to adapt models trained on adult English speech to better align with the target population (Danish speech from children), thus obtaining improved indicators of OCD. In this paper, we also weigh our methodological decisions.

Methods

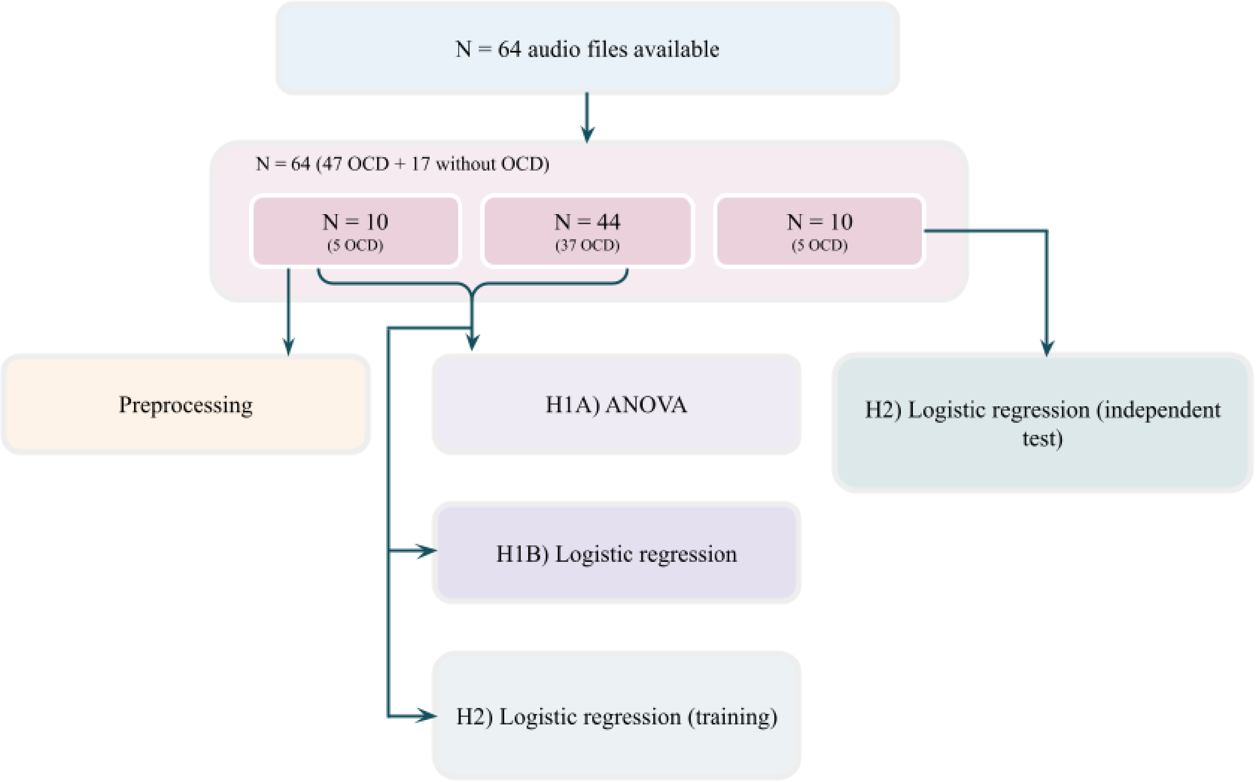
Participants and Setting

We will include audio recordings from a total of 64 youths aged 8 to 17 years (47 with OCD and 17 healthy controls). The audio recordings of diagnostic interviews stem from a large randomized clinical trial and case-control study of OCD, called TECTO [17]—TECTO runs at a public hospital in Denmark.

Ethics Approval

TECTO and the planned analyses have been approved by the ethics committee of the Capital Region of Denmark (H-18010607). The selection of participants in the current study is depicted in Figure 1.

Figure 1. Flow diagram of audio data selection. OCD: obsessive-compulsive disorder.


N = 64 audio files available
N = 64 (47 OCD + 17 without OCD)
N = 10
(5 OCD)
N = 44
(37 OCD)
N = 10
(5 OCD)
Preprocessing
H1A) ANOVA
H2) Logistic regression (independent test)
H1B) Logistic regression
H2) Logistic regression (training)

Measures

Trained mental health professionals conducted clinical interviews before inclusion in TECTO to establish diagnoses and OCD severity in patients and rule out psychiatric diagnoses in controls.

Diagnostic Status

Trained mental health professionals used a semistructured clinical interview—the Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS)—to screen for and establish diagnoses in participants [18]. The K-SADS is designed to establish psychiatric diagnoses in youth between the ages of 6 and 18 years.

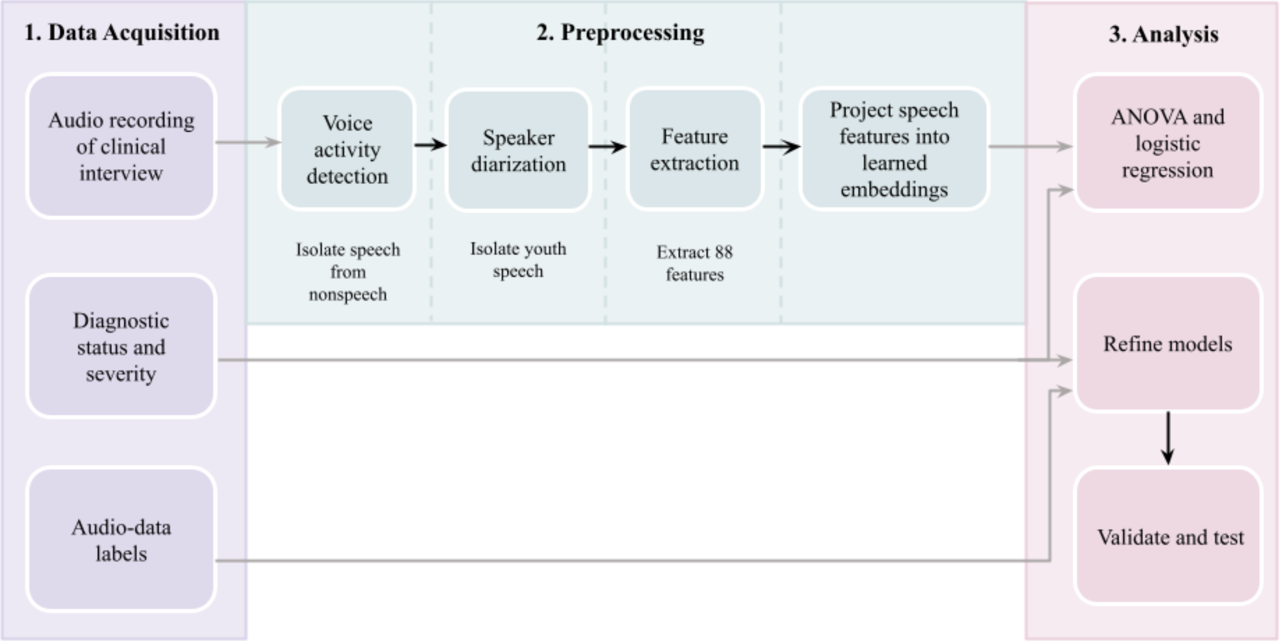
OCD Severity

Trained mental health professionals assessed the clinical severity of OCD using the Children’s Yale-Brown Obsessive Compulsive Scale (CY-BOCS) [19]. The CY-BOCS interview begins with a checklist of obsessions and compulsions to establish which symptoms have been present over the past week. On a scale from 0 to 4, clinicians rate 5 items on the severity of obsessions and 5 items on the severity of compulsions. Severity is rated on 5 dimensions: level of distress caused by the symptoms, functional interference, time consumed by symptoms, and resistance to and degree of control over symptoms. A total severity score is calculated by summing all 10 items on a scale from 0 to 40. Although CY-BOCS scores are continuous, a previous study of 815 youth between the ages of 4 and 18 years found the following cutoff scores to be consistent with global clinical severity ratings: 0-7, subclinical; 8-15, mild; 16-24, moderate; and 25-40, moderate-severe to severe [20]. A CY-BOCS score of ≥16 was required to be included in the study.

Audio Data Source

Samples of child speech will be taken from the K-SADS interviews. Owing to limited labeled data, we will use the first 10 minutes of the 30- to 90-minute interviews. Audio recordings were obtained using an on-camera microphone of a Sony video camera placed in various positions and in different rooms across observations. Data analysis is outlined in Figure 2.

Figure 2. Overview of the data analysis process.



Statistical Analysis Plan

The planned study has two main objectives:

1. Test the usefulness of a previously learned latent model, composed of 2 dimensions representing a compressed vocal feature space guided by activation labels [16], as a marker for OCD diagnosis and severity.
2. Learn a new latent model, which we will propose as an improved candidate indicator of OCD severity.

Our first objective will be achieved through 2 statistical analyses, with the following hypotheses:

1. Objective 1 (H1): an ANOVA will be conducted to test the first research hypothesis (objective H1A) that there is an effect of diagnosis (OCD vs no psychiatric diagnosis) on the vocal feature latent model. With logistic regression (H1B), we will test hypothesis H1, which states that there is an effect of the vocal feature latent model on OCD severity classifications in moderate to extreme OCD cases, with the vocal feature latent model corrected for age and gender. This analysis will only include patients with OCD, and we will examine the classification accuracy of this model.
2. Objective 2 (H2): the second objective will be achieved through data labeling and machine learning modeling. Here, outcomes will include diagnostic status (OCD vs no psychiatric diagnosis) and OCD severity (CY-BOCS severity scores). We will validate the modeling using a leave-two-individuals-out cross-validation and prediction accuracies. Subsequently, we will assess the obtained results using an independent test. We will extract data from 10 of the youths' audio samples post machine learning modeling for this purpose.

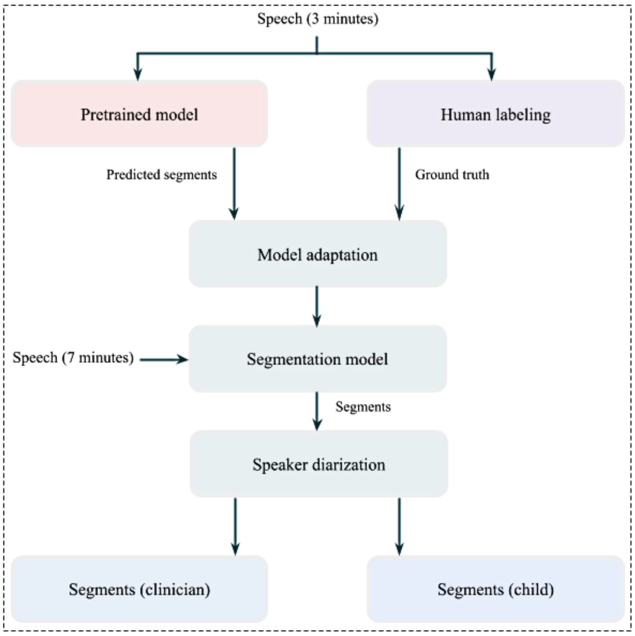
Processing Pipeline

Preprocessing Audio and Annotations

The audio preprocessing system is illustrated in Figure 3. To effectively use the audio signals, the audio recordings must be segmented into shorter segments. Next, speech and nonspeech regions must be differentiated. To obtain speaker segments, the conversations will first be segmented into speech and nonspeech regions.

We will use pretrained voice activation and diarization models that will be fine-tuned with a few manually annotated ground-truth labels of speech and nonspeech samples from our data set [21]. We follow this approach to ensure a balance between resources spent in training a model while maintaining high accuracy in the obtained speaker segments. Following speaker segmentation of the audio recordings, only the audio segments corresponding to the youth are retained.

Figure 3. Flow diagram of the preprocessing pipeline.



Pretrained Segmentation and Diarization Model

We used the vad-crdnn-libriparty pretrained model from SpeechBrain for speech detection [22] owing to its flexibility and modular structure, which is a favorable quality when applying models to clinical data owing to the tractability of errors. The pretrained model is based on a combination of convolutional and recurrent neural networks and a fully connected network. The model accepts audio segments as input and yields a posterior probability frame-level or segment-level posterior probability as output. Finally, a threshold is applied on the output posterior probability to classify the segments as speech and nonspeech.

We define a speaker segment to be an utterance of 2-4 seconds from a youth. The interviews will have a varying number of utterances. For a balanced analysis, we will randomly select 10 speaker segments per youth. Since the random selection of speaker segments may result in a high variance within the observations, we will also explore the feasibility of analyzing specific audio segments consistently over all speakers. These audio segments will be selected on the basis of insightful segments in the interview; for instance, questions in the interview associated with depression. In total, 10 interviews—5 with youth with OCD and 5 with those with no psychiatric diagnosis—have already been used to develop an appropriate method for speaker segmentation [21]. We will add the remaining 54 observations once we commence the described analysis.

Audio Features

From each speaker segment, we will extract the extended Geneva minimalistic acoustic parameter set [23], consisting of functionals of lower-level features. We will use the OpenSmile toolkit to extract these features and the resulting feature vector has a length of 88 [24].

The derived feature vector will be used as input to a neural network autoencoder model to obtain a latent model. The latent model is pretrained on English speech using the Interactive Emotional Dyadic Motion Capture data set and a semisupervised loss as previously described [16]. The latent model is a 2D latent space (*v*1 and *v*2) in a semisupervised denoising autoencoder with a reconstruction loss plus a loss based on the linear association between the activation labels and the latent space (the aim is a high association). Thus, to obtain *v*1 and *v*2, the 88 vocal features are projected into the latent space through the pretrained network. Previous analyses indicate that this latent space represents 2 dimensions of speech activation or intensity [16]. Sad and bored speech is characterized by low activation whereas elated and angry speech is characterized by high activation [25]. Thus, we assume that this latent space represents emotional intensity. We do not have an objective metric to evaluate this quantitatively using the current data and labels. Therefore, we will test this assumption in future work when more data labels are available.

For the machine learning analyses, we will use the raw speech signals to learn new features.

Statistical Models

Let *v*1 and *v*2 describe the vocal feature latent model adjusted for age and gender effects.

Statistical Models for Objective H1A

We will perform an ANOVA using a mixed effects regression model with the vocal features *v*1 and *v*2 as the outcome. We code the diagnosis, *Diagnosis*={0, 1} for no-OCD or OCD, respectively. The model for the *j*th vocal feature is as follows:

$$v_{ij} = \mu_j + \alpha_j(age_i) + \gamma_j(gender_{ii}) + v_j(age \times gender)_i + \delta_j(Diagnosis_i) + y_j(youth_i) + \varepsilon_{ij} \quad (1)$$

where μ_j is a constant offset for the *j*th endpoint, $i = 1, ..., 200$, $y_j\ (youth_i) \sim N\ (0,\ \sigma^2_{yj}\ )$, $\varepsilon_i \sim N\ (0,\ \sigma^2_j)$, and the effects are mutually independent. Age, gender, and an interaction between age and gender are included as fixed effects to remove confounding effects from these variables. Diagnosis is included as a fixed effect and is our primary interest. The individual youth, denoted *youth*, is included as a random effect. *Diagnosis* is included as a fixed effect. In total, 10 repeated measures are available for each of the 64 youths included in this model.

With this analysis, we shall test the hypothesis that there is no effect of diagnosis; that is, $H_0 : \delta_j = 0$, or equivalently that the means of the vocal features are equal for the OCD and no-OCD groups (when effects of age and gender are removed), at a 1.25% level of significance.

Statistical Models for Objective H1B

We will use a logistic regression with OCD severity categories moderate and severe-extreme as the outcomes. We denote the severity $s = 0, 1$. We will fit the following logistic regression model to the scores:

$$\text{logit(si)} = \theta + \text{b1}\,\overline{\boldsymbol{v}}\,\text{i} + \text{b2}\,\overline{\boldsymbol{v}}\,\text{i} + \varepsilon\text{i} \quad (2)$$

where $\varepsilon_i \sim N\,(0, \sigma_{2j})$ and v_j is the *j*th latent vocal feature adjusted for age and gender and averaged over the 10 repetitions for each youth, as described in the following.

This analysis aims to test the effects from vocal features on the severity scores through the null hypotheses that $b_1 = 0$ and $b_2 = 0$, which we will test at a 1.25% level of significance.

Adjusting for Confounding and Calculating Average Features

We expect age and gender to have an influence on the vocal features and would like to remove these confounding effects from the signals. We will exclude effects from age, gender, and age and gender interactions from the vocal features through a linear regression, as follows.

First, we will find the confounding effects using a linear regression model, which estimates age and gender contributions to vocal features as follows:

$$\hat{v}_{ji} = \hat{\mu}_j + \hat{\alpha}_j(age_i) + \hat{\gamma}_j(gender_i) + \hat{v}_j(age \times gender)_i \qquad (3)$$

where, for the *i*th observations and *j*th vocal feature, $j = 1, 2$. The vocal features adjusted for age and gender are then as follows:

$$\tilde{v}_j = v_j - \hat{v}_j \qquad (4)$$

Finally, we will use the average vocal features over the 10 repetitions for the *k*th youth to obtain $\overline{\boldsymbol{v}}_{jk} = \Sigma i\varepsilon youth_k\, v_{ji} / 10$.

Adjusting for Multiple Testing

We are interested in the null hypotheses associated with the effect of a diagnosis on the two dimensions in the latent model as well as the effects from the two adjusted latent dimensions on the severity scores. We will adjust our 4 *P* values using the Bonferroni method, and thus test each null hypothesis at a 5%/4=1.25% level of significance resulting in a family-wise error rate for the four tests of 5%.

Machine Learning Models for Objective 2

To improve the vocal features for possible use as biomarkers of OCD severity, we shall use the following methodology. First, 2 authors (SD and NLL) will annotate samples for activation and valence on a 5-point Likert scale with 1 indicating low activation and 5 indicating high activation. We shall only provide the annotators with one speech segment at a time in a blinded manner and in random order.

For the planned analyses, we will not use the valence labels, but the activation labels as these labels have shown promise in previous investigations [16]. We will use these labels as in our previous work [16]. The major advantages of labeling speech signals with activation scores include making labels for Danish speech and for children's speech, and labeling speech of patients and controls. Thus, the labels provide added information as no open source data exist, thus supporting transfer learning wherever feasible.

Subsequently, we shall train the following machine learning models using the following annotations: variational and denoising autoencoders with semisupervised losses based on activation labels and a reconstruction error, plus a logistic regression or logistic regression loss in a neural network to obtain a classification model.

Model Validation Strategies

We will use a leave-two-youth-out cross-validation, leaving 1 control plus 1 patient out for validation in each fold to tune the hyperparameters in our machine learning models and to choose between a simple logistic regression or a neural network with a logistic regression loss. To validate the results, we will use an independent test set of 10 interviews (5 with the patients and 5 with the controls) to evaluate the performance of these methods. These interviews will be transferred at the time of the independent test.

Software

Data will be processed and analyzed using the most current and reliable version of Praat [26], Audacity [27], Python [28], R (R Core Team) [29,30], and OpenSmile [24].

Results

Statistical Report Model 1A

The statistical analysis in this section was performed with simulated data, simulated with the same structure as described above; that is, repeated measures from 10 individuals and with age and gender effects on the vocal feature latent model. The model 1A residuals are illustrated in Figure 4.

No strong effects were found, indicating that the assumptions have been violated. Thus, in this case, we will analyze the results from the model. A summary of the fixed effects is provided in Table 1 and one for the random effects is given in Table 2.

Finally, the CIs for all the parameter estimates are summarized in Table 3. We shall repeat this analysis for *v*1 and *v*2. We shall compare the *P* value [$Pr(> |t|)$] from the diagnosis to the 1.25% level of significance to asses if we can reject our null hypothesis of no effect from diagnosis. In the simulation, 2.92×10^{-06} is less than .0125; thus, we would reject the null hypothesis.

Figure 4. Quantile-quantile plot of residuals from model 1A with end point v_j. Data are simulated and placeholders for results.

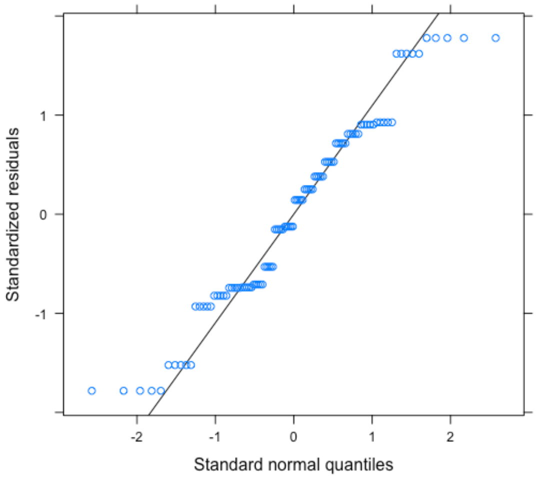


Table 1. Summary of fixed effects for v_j. Data are simulated and placeholder for results.

Effect	Estimate	SE	*t* test	*df*	*P* [r(>|*t*|)] value
Intercept (μ)	1.16	0.12	9.54	4.99	2.15×10^{-4}
Age (α)	0.49	0.01	51.23	4.99	5.36×10^{-8}
Gender (γ)	0.83	0.13	6.16	4.99	1.64×10^{-3}
Diagnosis (δ)	0.94	0.04	22.95	4.99	2.92×10^{-6}
Age:gender (ν)	0.02	0.01	1.19	4.99	2.87×10^{-1}

Table 2. Summary of simulated random effects v_j with 100 observations and 10 groups (youth).

Groups	Variance	SD
Youth (σ_y)	2.29×10^{-3}	4.79×10^{-2}
Residual (σ_e)	2.33×10^{-3}	4.83×10^{-2}

Table 3. Simulated CIs for effects on v_j.

Effect	Effects (%), CI	
	2.5%	97.5%
σ_y	1.87×10^{-2}	5.73×10^{-2}
σ_e	4.2×10^{-2}	5.63×10^{-2}
Intercept	9.71×10^{-1}	1.34
Age	4.76×10^{-1}	5.05×10^{-1}
Gender	6.23×10^{-1}	1.04
Diagnosis	8.74×10^{-1}	9.99×10^{-1}
Age:gender	-4.16×10^{-3}	3.34×10^{-2}

Statistical Report Model 1B

We shall first assess the assumptions of our model by examining the surrogate residuals in a quantile-quantile (q-q) plot (see Figure 5). In this case, the assumptions are violated as the residuals are not normally distributed. In case the assumptions are violated, we will not report on parameter estimates and their confidence intervals, as these will not be meaningful. However, we can still use the model for predictions.

Figure 6 illustrates the receiver operating characteristic (ROC) curve for the training data, and the estimated classification accuracy is 80%. Additionally, we will report sensitivity, specificity, area under the curve, and a confusion matrix, similar to the next section.

If the surrogate residuals show an approximated normal distribution in the q-q plot, we will report parameter estimates, etc, as we have for model 1A.

Figure 5. Quantile-quantile plot of surrogate residuals from model 1B with endpoint v_j. Data are simulated and placeholders for results.

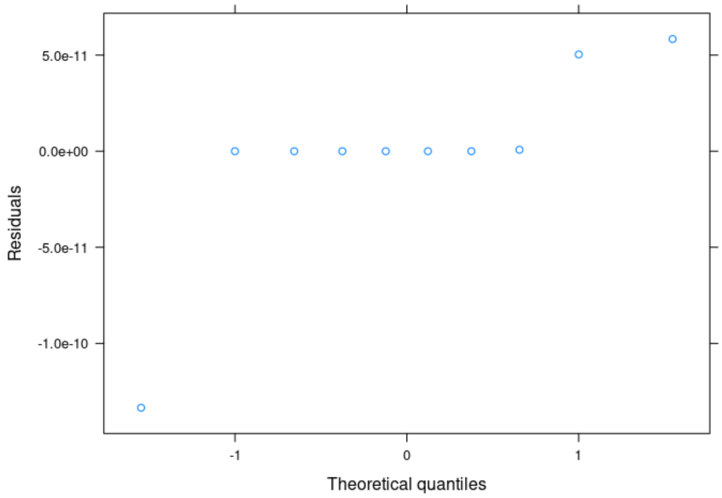
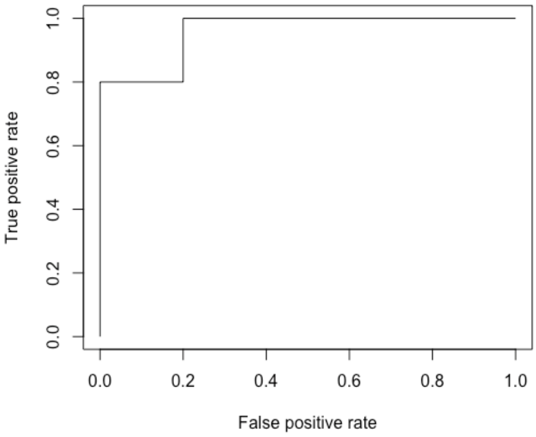


Figure 6. Receiver operating characteristic curve for the training data.



Statistical Report Model 2

The evaluation results of the machine learning model on simulated results are reported using the classification metrics in Table 4, the ROC curve in Figure 7 and the confusion matrix in Figure 8. The results will be presented for the validation data (the left-out youth) over the cross-validation to select hyperparameters and a model. We will also compare to the same measures for the training data to assess the amount of possible overfitting. Final results will be provided for the independent test data consisting of 10 youths.

Table 4. Simulated classification performance.

Evaluation metric	Value (normalized)
Accuracy	0.82
Sensitivity	0.78
Specificity	0.86
Area under the curve	0.89

Figure 7. Receiver operating characteristic curve for the classification model. AUC: area under the curve; OCD: obsessive-compulsive disorder.

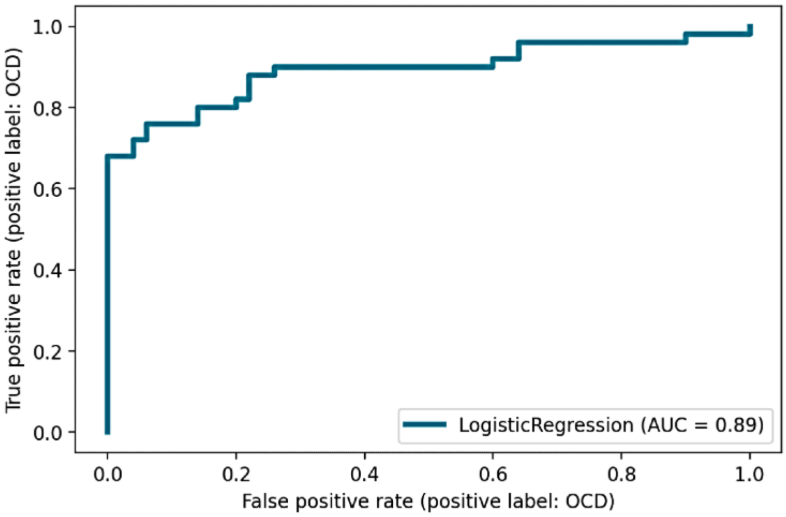
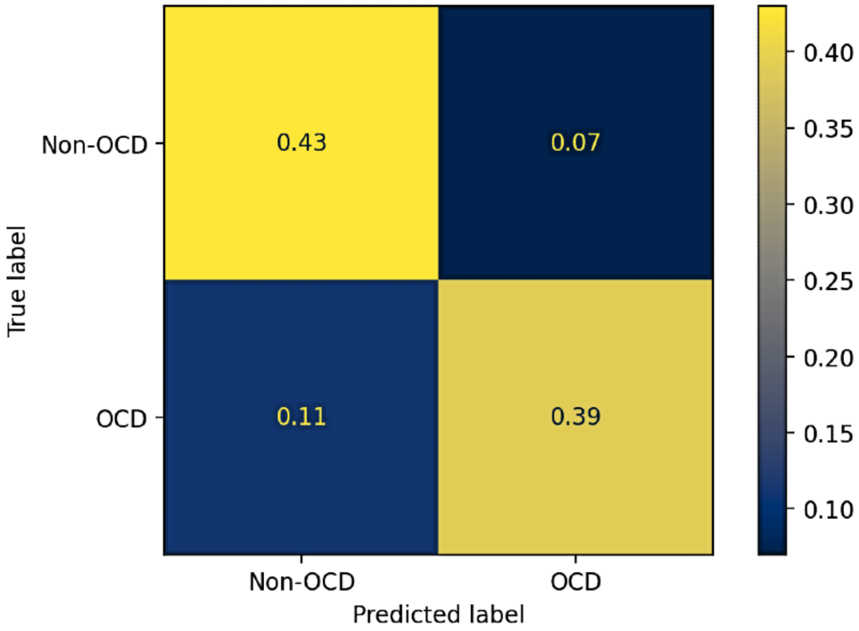


Figure 8. Confusion matrix for the classification model. OCD: obsessive-compulsive disorder.



Discussion

Expected Findings

In this paper, we describe the methods and analysis plan for testing the effects of OCD diagnosis and symptom severity on vocal features in the youth. We anticipate that we can improve the methods that model these putative associations. Through this work, we aim to obtain reliable transdiagnostic indicators of clinical severity from voice that would serve as valuable monitoring tools in psychiatry [31]. Additionally, the vocal indicators obtained from this work, in tandem with additional data modalities including video and semantics of the speech conversation, will be employed in multi-sensor modeling of OCD behavior in future work [32]. The results described in this analysis plan will be published in relevant scientific journals.

Strengths

A major strength of the planned study is this documentation of the analysis plan prior to performing analyses. Documented a priori hypotheses prevent unscientific practices such as selectively reporting significant results [33]. Furthermore, diagnostic status and OCD severity levels were established by trained mental health professionals using gold-standard clinical interviews. Another strength of the planned study is that we will include gender in the models. Compared with male speech, female speech is marked by a higher pitch. A study found that gender-specific models were more accurate in detecting sadness and positive and negative affect than gender-independent models [34]. Depressed speech is marked by reduced pitch, whereas anxious speech is marked by increased pitch [10].

Another study found that pitch appears more important for detecting stress in men while Root Mean Square Energy appears more influential in detecting stress in women [35]. Thus, if training data sets are overrepresented by one gender, an automatic depression or anxiety severity rating algorithm may misclassify speech by other genders. We also attempted to strengthen our analyses by adding age to the models. The available latent model that we will use in our study of youth of a wide age range were trained on adult speech. With age, pitch and formant frequencies tend to decrease [36]. From childhood to adolescence, speech rate increases, and conveying emotions with prosodic cues, including pitch and timing, is a development feat [37].

Limitations

The planned analyses have some foreseeable challenges. First, we plan to recruit a small sample. Preprocessing and data labeling, required for the planned analyses, are time-consuming tasks. Our small sample size increases the risk of type II errors. Machine learning techniques are more robust as we can improve models by adding more labeled data and testing on a new independent data set. Second, the audio samples that will be analyzed were collected for clinical purposes and not under optimal conditions for audio feature extraction and analysis (ie, using high quality microphones in a quiet environment). Thus, methods developed on these data will likely translate well to other naturalistic settings [38-41]. Finally, this study focuses on the associations among OCD diagnosis, severity, and vocal features in the child. In future work, we will study the effect of OCD diagnosis on vocal features while accounting for secondary diagnoses and investigate the influence of clinician's vocal characteristics on the vocal features of the child.

Conclusions

This predefined plan will limit bias in the interpretations and conclusions of the reported results of the future publication. If the results in the planned study are promising, this will be a step toward using vocal sensing to automate objective assessments and monitoring of severity of psychiatric disorders such as OCD.

Acknowledgments

We would like to thank Sofie Heidenheim Christensen, the TECTO project coordinator, for her part in recruiting participants and other ways in which she managed aspects of the project that made this work possible. This work is funded by a Novo Nordisk Foundation grant (NNF19OC0056795).

Data Availability

We are not permitted to share voice samples or other personal identifying information. We will investigate the possibility of making an anonymized data set with extracted voice features available. If possible, we shall indicate where to obtain the data set with the published results.

Conflicts of Interest

None declared.

References

1. Meier SM, Mattheisen M, Mors O, Schendel DE, Mortensen PB, Plessen KJ. Mortality among persons with obsessive-compulsive disorder in Denmark. JAMA Psychiatry 2016 Mar;73(3):268-274 [FREE Full text] [doi: 10.1001/jamapsychiatry.2015.3105] [Medline: 26818216]
2. Piacentini J, Bergman RL, Keller M, McCracken J. Functional impairment in children and adolescents with obsessive-compulsive disorder. J Child Adolesc Psychopharmacol 2003;13 Suppl 1:S61-S69. [doi: 10.1089/104454603322126359] [Medline: 12880501]
3. OCD Clinical Practice Review Task Force. Clinical Practice Review for OCD: Diagnostic and Statistical Manual of Mental Disorders (DSM-5). Anxiety & Depression Association of America. 2015 Jun 15. URL: https://adaa.org/resources-professionals/practice-guidelines-ocd [accessed 2022-09-12]
4. Obsessive-compulsive disorder and body dysmorphic disorder: treatment. National Institute for Health and Care Excellence. 2005 Nov 29. URL: https://www.nice.org.uk/guidance/cg31 [accessed 2022-09-12]
5. Benito KG, Walther M. Therapeutic process during exposure: habituation model. J Obsessive Compuls Relat Disord 2015 Jul 01;6:147-157 [FREE Full text] [doi: 10.1016/j.jocrd.2015.01.006] [Medline: 26258012]
6. Skarphedinsson G, De Nadai AS, Storch EA, Lewin AB, Ivarsson T. Defining cognitive-behavior therapy response and remission in pediatric OCD: a signal detection analysis of the Children's Yale-Brown Obsessive Compulsive Scale. Eur Child Adolesc Psychiatry 2017 Jan;26(1):47-55 [FREE Full text] [doi: 10.1007/s00787-016-0863-0] [Medline: 27209422]
7. Benito KG, Machan J, Freeman JB, Garcia AM, Walther M, Frank H, et al. Measuring fear change within exposures: functionally-defined habituation predicts outcome in three randomized controlled trials for pediatric OCD. J Consult Clin Psychol 2018 Jul;86(7):615-630 [FREE Full text] [doi: 10.1037/ccp0000315] [Medline: 29939055]
8. Kraepelin E. Manic depressive insanity and paranoia. J Nerv Ment Dis 1921;53(4):350. [doi: 10.1097/00005053-192104000-00057]
9. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. Speech Communication 2015 Jul;71:10-49. [doi: 10.1016/j.specom.2015.03.004]
10. Low DM, Bentley KH, Ghosh SS. Automated assessment of psychiatric disorders using speech: a systematic review. Laryngoscope Investig Otolaryngol 2020 Feb;5(1):96-116 [FREE Full text] [doi: 10.1002/lio2.354] [Medline: 32128436]

11. Cassol M, Reppold CT, Ferrão Y, Gurgel LG, Almada CP. Análise de características vocais e de aspectos psicológicos em indivíduos com transtorno obsessivo-compulsivo. Rev soc bras fonoaudiol 2010 Dec;15(4):491-496. [doi: 10.1590/s1516-80342010000400004]
12. Garcia-Real T, Diaz-Roman TM, Garcia-Martinez V, Vieiro-Iglesias P. Clinical and acoustic vocal profile in children with attention deficit hyperactivity disorder. J Voice 2013 Nov;27(6):787.e11-787.e18. [doi: 10.1016/j.jvoice.2013.06.013] [Medline: 24246332]
13. Adams P, Rabbi M, Rahman T, Matthews M, Voida A, Gay G, et al. Towards Personal Stress Informatics: Comparing Minimally Invasive Techniques for Measuring Daily Stress in the Wild. 2014 Presented at: 8th International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2014; Oldenburg. [doi: 10.4108/icst.pervasivehealth.2014.254959]
14. Holmqvist-Jämsén S, Johansson A, Santtila P, Westberg L, von der Pahlen B, Simberg S. Investigating the role of salivary cortisol on vocal symptoms. J Speech Lang Hear Res 2017 Oct 17;60(10):2781-2791. [doi: 10.1044/2017_JSLHR-S-16-0058] [Medline: 28915296]
15. McGinnis EW, Anderau SP, Hruschak J, Gurchiek RD, Lopez-Duran NL, Fitzgerald K, et al. Giving voice to vulnerable children: machine learning analysis of speech detects anxiety and depression in early childhood. IEEE J Biomed Health Inform 2019 Nov;23(6):2294-2301 [FREE Full text] [doi: 10.1109/JBHI.2019.2913590] [Medline: 31034426]
16. Das S, Lund NL, Lønfeldt NN, Pagsberg AK, Clemmensen LH. Continuous metric learning for transferable speech emotion recognition and embedding across low-resource languages. In: Proceedings of the Northern Lights Deep Learning Workshop 2022. Tromsø: Septentrio Academic Publishing; Apr 06, 2022.
17. Pagsberg AK, Uhre C, Uhre V, Pretzmann L, Christensen SH, Thoustrup C, et al. Family-based cognitive behavioural therapy versus family-based relaxation therapy for obsessive-compulsive disorder in children and adolescents: protocol for a randomised clinical trial (the TECTO trial). BMC Psychiatry 2022 Mar 19;22(1):204 [FREE Full text] [doi: 10.1186/s12888-021-03669-2] [Medline: 35305587]
18. Puig-Antich J, Ryan N. Kiddie schedule for affective disorders and schizophrenia. Pittsburgh, PA: Western Psychiatric Institute; 1986.
19. Scahill L, Riddle MA, McSwiggin-Hardin M, Ort SI, King RA, Goodman WK, et al. Children's Yale-Brown Obsessive Compulsive Scale: reliability and validity. J Am Acad Child Adolesc Psychiatry 1997 Jun;36(6):844-852 [FREE Full text] [doi: 10.1097/00004583-199706000-00023] [Medline: 9183141]
20. Lewin AB, Piacentini J, De Nadai AS, Jones AM, Peris TS, Geffken GR, et al. Defining clinical severity in pediatric obsessive-compulsive disorder. Psychol Assess 2014 Jun;26(2):679-684. [doi: 10.1037/a0035174] [Medline: 24320764]
21. Das S, Lønfeldt N, Pagsberg A, Clemmensen L. Speech detection for child-clinician conversations in Danish for low-resource in-the-wild conditions: a case study. arXiv Preprint posted online April 25, 2022. [doi: 10.48550/arXiv.2204.11550]
22. Ravanelli M, Parcollet T, Plantinga P, Rouhe A, Cornell S, Lugosch L, et al. SpeechBrain: a general-purpose speech toolkit. arXiv Preprint posted online June 8, 2021. [doi: 10.48550/arXiv.2106.04624]
23. Eyben F, Scherer KR, Schuller BW, Sundberg J, Andre E, Busso C, et al. The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for voice research and affective computing. IEEE Trans Affective Comput 2016 Apr 1;7(2):190-202. [doi: 10.1109/taffc.2015.2457417]
24. Eyben F, Wöllmer M, Schuller B. Opensmile: the munich versatile and fast open-source audio feature extractor. 2010 Presented at: MM '10: ACM Multimedia Conference; October 25-29, 2010; Firenze. [doi: 10.1145/1873951.1874246]
25. Alonso JB, Cabrera J, Medina M, Travieso CM. New approach in quantification of emotional intensity from the speech signal: emotional temperature. Expert Syst Appl 2015 Dec;42(24):9554-9564. [doi: 10.1016/j.eswa.2015.07.062]
26. Boersma P, Weenink D. Praat: doing phonetics by computer. URL: https://www.fon.hum.uva.nl/praat/ [accessed 2022-09-12]
27. Audacity. URL: https://audacityteam.org/ [accessed 2022-09-12]
28. Van Rossum G, Drake Jr JF. Python tutorial. Centrum voor Wiskunde en Informatica Amsterdam. 1995. URL: https://www.google.com/ [accessed 2022-09-12]
29. The R Project for Statistical Computing. URL: https://www.r-project.org/ [accessed 2022-09-12]
30. Liu Q, Shepherd B, Li C. PResiduals: an R Package for residual analysis using probability-scale residuals. J Stat Soft 2020;94(12):1-27. [doi: 10.18637/jss.v094.i12]
31. Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? Mol Psychiatry 2012 Dec;17(12):1174-1179. [doi: 10.1038/mp.2012.105] [Medline: 22869033]
32. Lønfeldt N, Frumosu F, Lund N, Das S, Pagsberg A, Clemmensen L. Computational behavior recognition in child and adolescent psychiatry: a statistical and machine learning analysis plan. arXiv Preprint posted online May 11, 2022.
33. Forstmeier W, Wagenmakers E, Parker TH. Detecting and avoiding likely false-positive findings - a practical guide. Biol Rev Camb Philos Soc 2017 Nov;92(4):1941-1968. [doi: 10.1111/brv.12315] [Medline: 27879038]
34. Black MP, Katsamanis A, Baucom BR, Lee C, Lammert AC, Christensen A, et al. Toward automating a human behavioral coding system for married couples' interactions using speech acoustic features. Speech Communication 2013 Jan;55(1):1-21. [doi: 10.1016/j.specom.2011.12.003]
35. Zuo X, Fung PN. A cross gender and cross lingual study on acoustic features for stress recognition in speech. 2011 Presented at: 17th International Congress of Phonetic Sciences; August 17-21, 2011; Hong Kong.

36. Gautam S, Singh L. The development of spectral features in the speech of Indian children. Sādhanā 2019 Jan 2;44(1). [doi: 10.1007/s12046-018-1028-2]
37. Dilley LC, Wieland EA, Gamache JL, McAuley JD, Redford MA. Age-related changes to spectral voice characteristics affect judgments of prosodic, segmental, and talker attributes for child and adult speech. J Speech Lang Hear Res 2013 Feb;56(1):159-177 [FREE Full text] [doi: 10.1044/1092-4388(2012/11-0199)] [Medline: 23275414]
38. Cummins N, Sethu V, Epps J, Schnieder S, Krajewski J. Analysis of acoustic space variability in speech affected by depression. Speech Communication 2015 Dec;75:27-49. [doi: 10.1016/j.specom.2015.09.003]
39. Alghowinem S, Goecke R, Epps J, Wagner M, Cohn J. Cross-cultural depression recognition from vocal biomarkers. Proc. Interspeech 2016:1943-1947. [doi: 10.21437/Interspeech.2016-1339]
40. Das S, Lønfeldt N, Pagsberg A, Clemmensen L. Towards transferable speech emotion representation: on loss functions for cross-lingual latent representations. 2022 Presented at: ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); May 23-27, 2022; Singapore. [doi: 10.1109/icassp43922.2022.9746450]
41. Das S, Lund N, Lønfeldt N, Pagsberg A, Clemmensen L. Zero-shot Cross-lingual Speech Emotion Recognition: A Study of Loss Functions and Feature Importance. 2022 Presented at: ISCA Symposium on Security and Privacy in Speech Communication 2022 (forthcoming); 2022; Incheon, Korea.

Abbreviations

CBT: cognitive behavioral therapy
CY-BOCS: Children’s Yale-Brown Obsessive Compulsive Scale
ERP: exposure and response prevention
K-SADS: Kiddie Schedule for Affective Disorders and Schizophrenia
OCD: obsessive-compulsive disorder
q-q: quantile-quantile
ROC: receiver operating characteristic

Edited by T Leung; submitted 30.05.22; peer-reviewed by E Casiraghi, B Johnson; comments to author 20.07.22; revised version received 27.07.22; accepted 30.07.22; published 28.10.22

Please cite as:
Clemmensen LKH, Lønfeldt NN, Das S, Lund NL, Uhre VF, Mora-Jensen ARC, Pretzmann L, Uhre CF, Ritter M, Korsbjerg NLJ, Hagstrøm J, Thoustrup CL, Clemmesen IT, Plessen KJ, Pagsberg AK
Associations Between the Severity of Obsessive-Compulsive Disorder and Vocal Features in Children and Adolescents: Protocol for a Statistical and Machine Learning Analysis
JMIR Res Protoc 2022;11(10):e39613
URL: https://www.researchprotocols.org/2022/10/e39613
doi: 10.2196/39613
PMID: