Voices of change: Associations between vocal markers and symptoms of ADHD - Findings from the LIFE Child Study

Rachel Bamberger

r.bamberger@fz-juelich.de

Research Centre Jülich

Leon D. Lotter
Max Planck School of Cognition

Nicolás Nieto
Research Centre Jülich

Tanja Poulain
Leipzig University

Antje Körner
Leipzig University

Wieland Kiess
Leipzig University

Michael Fuchs
University Hospital Leipzig

Georg von Polier
University Hospital Leipzig

Abstract

In the context of precision psychiatry, voice characteristics in psychiatric disorders have significant potential as diagnostic markers. However, most studies have focused on adults, resulting in a poor understanding of paediatric voice characteristics. The diagnosis of attention-deficit/hyperactivity disorder (ADHD) lacks objectivity and could potentially benefit from the inclusion of voice features as objective markers. This study used linear mixed models and machine learning to investigate associations between voice characteristics and parent-reported symptoms of hyperactivity and inattention in a population-based cohort. Two tasks were used to derive speaking-voice characteristics: counting and sustaining a tone. 2418 recordings from 1460 children of the LIFE Child cohort aged 5–18 years (49% girls) were included. We analysed demographic and health data, voice parameters, and hyperactivity/inattention scores from the Strengths and Difficulties Questionnaire (SDQ). Results indicated linear associations between voice features, particularly fundamental frequency and voice intensity, and SDQ scores. Machine learning models predicted hyperactivity/inattention scores with moderate accuracy (r = 0.36). Predictive performance was higher in girls than in boys, an effect largely attributable to overall lower symptom ratings by parents in girls. Our findings support the inclusion of voice characteristics in research on diagnostic markers in paediatric populations and underline the need for further studies to refine analytic approaches.

1 Introduction

At a time when around one in four children experience mental health problems [1, 2], *precision psychiatry* is gaining attention as promising approach to personalised care. The use of preventive, diagnostic, and therapeutic methods adapted to individual patients' characteristics has been studied for more than two decades [3]. Precision psychiatry aims to individualise mental health care e.g. by increasing accuracy of diagnosis and treatment [4]. In the context of child and adolescent psychiatry, efforts have focused on developing more accurate and precise diagnostic or risk stratification tools to tackle current, largely subjective, diagnostic practice [5]. One approach to identifying objective markers is breaking down disorders into their underlying pathophysiological features [6], which can be detected through measuring biosignals [7].

Acoustic biosignals are gaining recognition as tools to study pathophysiological mechanisms of psychiatric disorders [8, 9, 10]. Voice features may contain essential information about an individual's state and potentially indicate pathologies [7, 11]. Human voice is produced by complex interactions between the cognitive and neuromuscular system, making voice a highly sensitive output system [7]. Current research in adults indicates decreases in the fundamental frequency (f0, the basic pitch of the voice) and f0 range in individuals with major depressive disorder, which can result in monotonous intonation and a deeper voice [12]. In anxiety disorders, studies have found a significant increase in the f0 [12]. A recent scoping review points to the potential of paediatric voice as a promising method for the detection of several childhood conditions and disorders, such as autism spectrum disorder [13]. If voice features were translated into health-relevant indices, they could eventually become part of a more objective assessment within *biomarker batteries* [11].

One promising field of voice related biomarker research is attention-deficit/hyperactivity disorder (ADHD): Machine learning (ML) based studies in adults with ADHD have demonstrated the predictive utility of voice

features for ADHD classification, as well as associations with symptom severity [14]. Biologically, ADHD is linked to changes in the brain's dopamine signals, which impact motor control [15]. Voice production as well is linked to dopamine levels and involves complex motor behaviour. Thus, changes in dopamine pathways related to ADHD may alter vocal patterns in people with ADHD [16]. Children with ADHD seem to show altered voice features (e.g., f0, intensity, jitter) and more hyperfunctional vocal behaviour (e.g., more shouting; excessive, louder, faster talking) in comparison to their peers [17, 18]. These bodily changes make children with ADHD prone to voice disorders [19, 20]. Investigating this component of ADHD pathophysiology, in the form of voice features, could provide new insights and enrich the current state of knowledge in child and adolescent precision psychiatry [5, 21].

Our population-based study investigated children's voice features that may serve as indices of ADHD using multivariate statistics, including linear mixed models (LMM) and machine learning (ML) approaches. We aimed to evaluate the association of different acoustic parameters obtained under several vocal tasks with symptoms of *hyperactivity/inattention* (HI), as measured by the HI subscale of the *Strengths and Difficulties Questionnaire* (SDQ) [22] in a large epidemiological cohort. Based on previous findings [17, 18], we hypothesized that symptoms of hyperactivity/inattention are associated with higher voice intensity. Moreover, we explored associations of further voice features including fundamental frequency (f0), jitter, maximum phonation time (MPT) and *Dysphonia Severity Index* (DSI) and HI symptoms based on preliminary findings from earlier research in small samples [17, 18, 23].

2 Participants and methods

2.1 Study design and participants

This study analysed data from the *LIFE Child study*. This is an ongoing epidemiological longitudinal cohort study conducted at the *Research Centre for Civilization Diseases* in Leipzig, Germany (for details see [24]). The LIFE Child study protocols adhere to the Declaration of Helsinki. Written informed consent was received from all parents and children over 12 years of age.

Inclusion criteria for our study were children aged between 5 and 18 years, for whom both speaking-voice examination and information on SDQ HI values were available for one or more time points. Data from 1460 children (female: $n = 716$; male: $n = 744$) with a mean age of 10.96 years (range: 5.65–18.22) were included. Participants were either part of the *Health Cohort* ($n = 1297$) or the *Obesity Cohort* ($n = 163$). Examinations took place between the years 2012 and 2015. The number of visits varied between 1 ($n = 736$), 2 ($n = 462$), 3 ($n = 231$), or 4 ($n = 31$).

2.2 Methods

2.2.1 Demographic and health data

Demographic and health data were collected either by clinical interviews/questionnaires or clinical examination. All examinations were carried out by trained medical investigators/paediatricians in the child-friendly facilities of the outpatient clinic of the *Research Centre for Civilisation Diseases* in Leipzig. In addition to voice-specific data, child's age, sex, *Body Mass Index* (BMI), socio-economic status (SES) and pubertal

status were included in our study. Pubertal status was included due to missing information on voice change in most participants. The examination methods relevant to this study are described below and in detail in the study protocol [24] .

Socio-economic status

SES was calculated using a multidimensional index based on information provided by parents on their school education and vocational training, their occupational status, and their net equivalent income. The operationalisation used is based on the *KiGGS* study of the *Robert Koch Institute* [25]. SES scores can range from 3.0 to 21.0, with values indicating low (3.0–8.4), medium (8.5–15.4), or high (15.5–21.0) SES.

BMI-standard deviation score (BMI-SDS)

The original BMI value was calculated using the body mass (kilogram) divided by the square of the body height (m^2) [24]. To ensure comparability between age and sex groups, BMI-SDS (standard-deviation-scores) were calculated using the Kromeyer-Hauschild reference [26].

Pubertal status

The assessment of pubertal status as an indicator of physical maturity is based on Tanner's criteria for pubertal development [27, 28]. For this study, a corrected form of the pubertal status was used. In some cases, Tanner's criteria did not align with corresponding hormone levels (endocrine values). Certain endocrine values indicated a lower or higher physical maturity. To address this discrepancy, laboratory data were used to adjust the Tanner stage where necessary. Values range from 1–5, with higher values indicating greater maturity.

2.2.2 Voice measurements and features

Voice measurements were conducted according to a standard operating procedure based on the recommendations by the *Union of the European Phoniatrics* [29] and as described by Berger et al. [30]. Voice examination was carried out in a soundproof room by trained investigators. Recording and analysis of the voice was done using the *DiVas Software* (XION medical, Berlin, Germany) and a self-calibrating USB-microphone-headset that was kept at a constant distance of 30 cm from the children's mouth. Participants had two different tasks. For the analysis of the speaking voice, the task was to count from 21 to 30 at five different sound pressure levels: quietest voice (quiet_I), conversational voice (conversation_II), presentation voice (presentation_III), loudest voice (loud_IV), and quietest voice again (quiet_V) for a voice reset test. In different conditions, children were instructed to speak as soft as possible, but to avoid whispering (I); to speak as if in a normal conversation sitting across from each other (II); to speak as if in a classroom giving a presentation (III), or to shout with their loudest voice, but without screaming (IV). Immediately after the shouting task, the children were instructed to count as quietly as possible again (V). This voice reset test is considered unremarkable if the values are comparable to those of the initial quiet measurement [30, 31]. The second task consisted of sustaining a tone on "na" at a comfortable, medium pitch and volume. After a maximal inhalation, the tone should be sustained for as long as possible. Afterwards, the tone should be sustained once more for a few seconds, aiming for the sound to be as clear and pure as possible.

For the analysis, we included the following parameters: fundamental frequency (f0; Hz) and intensity (dB(A)), extracted from each condition of the first task and maximum phonation time (MPT) and Jitter, extracted from the second task. Additionally, the highest pitch (f0max) and the lowest sound pressure level (SPLmin) achievable with the singing voice were measured. The Dysphonia Severity Index (DSI) was calculated from f0max, SPLmin, MPT, and jitter using a formula by Wuyst et al. [32]. The DSI ranges from + 5 (normal voice) to − 5 (severely dysphonic voice). While normative DSI values are well established for adults [33], there are fewer normative data available for children. Since children's voices change due to growth and puberty (voice change), DSI values in children are highly variable and age dependent. Compared to adults, children often show slightly lower DSI values. A total of 13 voice-derived features were used for analysis.

2.2.3 Behavioural difficulties/strengths

Behavioural data was extracted from the subscale *hyperactivity/inattention* (HI) from the parent-version of the SDQ, a widely used instrument for measuring behaviour in children and adolescents [22]. The subscale HI consists of five questions regarding hyperactivity/impulsivity (e.g. restless/overactive; fidgeting) and inattention (e.g. easily distracted/concentration wanders) on a three-point Likert scale, resulting in scores ranging from zero to ten.

2.4 Statistical Analysis

For statistical analysis we used two approaches. Linear mixed models were used for exploratory analysis and to examine associations between variables. Machine learning was applied to assess the feasibility of making predictions at the individual level. All analyses were conducted using *Python* (3.9), leveraging libraries such as *scikit-learn* [34]. A complete list of libraries and their versions can be found in the public GitHub repository.

2.4.1 Linear mixed models

LMM analysis preprocessing included normalising the data by z-scaling all variables. For the main analysis, voice features were set as independent variables, SDQ HI scores were set as dependent variables, and covariates included age, sex, SES, and BMI-SDS. Additionally, we conducted sex-stratified analyses and a robustness analysis including pubertal status as additional covariate. Fixed effects included the main predictors of interest, i.e. voice features and covariates. To account for the repeated measurements, random effects included a grouping variable (individual ID of the child to control for within-subject correlations) and a random slope for the variable *number of visits* was added. LMMs were fitted using *statsmodels* and the *Powell* method was used to optimise the likelihood function (1000 iterations). Models were evaluated by calculating the variance of the fixed effects and residuals, reported as coefficients, standard errors, confidence intervals, and p-values. The Benjamini-Hochberg procedure was applied to control for the False Discovery Rate (FDR). Results with FDR-corrected p-values (q-values) < 0.05 were considered statistically significant.

2.4.2 Machine learning

We conducted a series of ML analyses using *Ridge Regression*, that differed in the covariates or data used. The main analysis included voice features along with age, sex, SES, BMI-SDS as covariates. To assess the added predictive value of the voice features, we compared this model to one that included only the covariates.

The difference in the MAEs was evaluated using the Nadeau and Bengio's corrected t-test [35]. Additionally, we trained a model using only voice features. We used permutation testing in all models to determine whether the model's performance was significantly better than chance. Finally, to analyse effects of sex and pubertal status, we conducted a robustness check.

Prior to model fitting, all numerical features were standardized. When applicable, covariates were regressed out of the voice features using linear confound removal. Hyperparameter tuning was performed using *grid search*, evaluating three feature set sizes (top 5, 10, or all features) and six Ridge regression α values ranging from 0.0001 to 10.0 on a logarithmic scale. Model evaluation included 20 iterations of a 5-fold nested cross-validation. *GroupKFold* was used to ensure that repeated measurements from the same individual did not appear in both training and test sets. Outer folds assessed model performance while inner folds handled hyperparameter tuning. MAE, R^2, and *r* were calculated to evaluate performance. Statistical significance was assessed via permutation testing, comparing observed performance to a null distribution obtained by randomly shuffling target labels. Results with $p < 0.05$ were considered statistically significant. Finally, we calculated *Shapley Additive Explanation* (SHAP) values to identify which features had the greatest influence on model predictions.

3 Results

3.1 Descriptive statistics

Table 1 shows baseline study, demographic, and health characteristics from all included participants. Descriptive statistics on voice features and the SDQ HI from all visits can be found in **Table 2**. The exact sample sizes and number of features used for each model are in the LMM and ML Results sections, as these varied with the model specifications.

Table 1 Baseline study, demographic and health characteristics

Characteristic	n (%)	Mean/n (SD/%)
Sample	1460 (100)	1460 (100%)
Health cohort		1297 (88.8%)
Obesity cohort		163 (11.2%)
Number of visits	1460 (100)	1.70 (0.81)
1		736 (50.4%)
2		462 (31.6%)
3		231 (15.8%)
4		31 (2.1%)
Age (years)	1460 (100)	10.96 (3.02)
Sex	1460 (100)	
Female		716 (49.0%)
Male		744 (51.0%)
SES	1402 (96.03)	
Low		196 (14.0%)
Middle		856 (61.1%)
High		350 (24.9%)
BMI-SDS	1449 (99.25)	0.24 (1.23)
Health cohort		-0.03 (0.99)
Obesity cohort		2.41 (0.62)
Pubertal status	1205 (82.53)	2.20 (1.52)

Note. SES = socio-economic status; BMI-SDS = BMI-standard deviation score

Table 2 Characteristics of voice and SDQ HI variables (all visits)

Characteristic	n	Mean (SD)	Range
SDQ			
HI	2418	3.06 (2.47)	0.00−10.00
Voice features			
f0_quiet_I	2418	186.67 (43.63)	80.39−347.03
f0_conversation_II	2418	204.39 (43.62)	89.04−367.50
f0_presentation_III	2417	227.59 (47.56)	87.99−374.36
f0_loud_IV	2411	296.60 (55.73)	124.24−425.38
f0_quiet_V	2370	198.23 (46.78)	90.30−371.00
spl_ quiet_I	2418	51.46 (3.17)	41.62−65.14
spl_ conversation_II	2418	58.93 (3.29)	48.56−73.56
spl_ presentation_III	2417	65.88 (4.16)	54.41−86.86
spl_ loud_IV	2411	81.47 (6.30)	48.45−102.82
spl_ quiet_V	2370	54.06 (3.53)	43.85−82.51
Jitter	2357	1.89 (3.49)	0.13−42.65
MPT	2357	13.05 (5.01)	1.59−36.32
DSI	2357	1.77 (4.42)	-46.74−10.03

Note. SDQ = Strengths and Difficulties Questionnaire; HI = subscale hyperactivity/inattention; f0 = fundamental frequency (Hz); spl = sound pressure level (intensity; dB(A); MPT = maximum phonation time; DSI = dysphonia-severity-index

3.2 Correlation analysis: relevant covariates

Covariates included age, sex, SES, BMI-SDS and pubertal status. They were selected based on previously reported associations with the SDQ [36, 37] or voice features [38], and the results of a Spearman-Rho correlation analysis (see **Figure 1**). To handle multicollinearity, highly correlated predictors ($r > 0.8$) were not analysed in the same LMM model; therefore, age and pubertal status were analysed in separate models.

3.3 LMM

3.3.1 LMM: Main analysis

The main LMM analysis revealed that several voice variables were significantly associated with SDQ HI values (**Table 3**). The total number of included values per voice feature (n) varied between 2370 and 2418, due to

data availability for the different conditions. Including age, sex, SES and BMI-SDS as covariates, and correcting for multiple testing, fundamental frequency (f0) in conditions 2–5 and voice intensity (spl) in all conditions were found to be significantly associated with HI values. These associations indicated that higher SDQ HI values corresponded with a voice that was used louder and at a higher pitch during the counting task. The largest effects were detected in the *voice reset test* (quiet_V, speaking quiet again) for both f0 (β = 0.075, q < 0.001) and intensity (β = 0.091, q < 0.001) (**Figure 2**). No significant effects were found for other voice features, namely Jitter (β = 0.024, q = 0.104), MPT (β = -0.019, q = 0.332), and DSI (β = -0.022, q = 0.152).

Table 3 Significant associations between SDQ HI and voice variables (linear mixed model)

SDQ	Voice feature	*n*	df_resid	coef (β)	se	ci_lower	ci_upper	p	q
HI	f0_conversation_II	2418	2412	0.070	0.019	0.033	0.106	< 0.001	< 0.001
	f0_presentation_III	2417	2411	0.057	0.020	0.018	0.096	0.005	0.009
	f0_loud_IV	2411	2405	0.037	0.016	0.006	0.068	0.021	0.029
	f0_quiet_V	2370	2364	0.075	0.017	0.041	0.109	< 0.001	< 0.001
	spl_ quiet_I	2418	2412	0.070	0.016	0.038	0.101	< 0.001	< 0.001
	spl_ conversation_II	2418	2412	0.054	0.016	0.023	0.086	0.001	0.002
	spl_ presentation_III	2417	2411	0.037	0.016	0.006	0.068	0.018	0.026
	spl_ loud_IV	2411	2405	0.049	0.016	0.017	0.081	0.002	0.005
	spl_ quiet_V	2370	2364	0.091	0.016	0.059	0.123	< 0.001	< 0.001

Note. All models were adjusted for the covariates age, sex, SES, BMI-SDS; all models converged successfully; degrees of freedom for all models are 9; SDQ HI = Strengths and Difficulties Questionnaire subscale hyperactivity/inattention; f0 = fundamental frequency; spl = sound pressure level (intensity); df_resid = degrees of freedom (residual); coef = coefficient; se = standard error; ci_lower/upper = confidence interval (lower/upper bound); p/q = p-/q-value (FDR adjusted p-value)

The robustness analysis, which included puberty status instead of age, indicated stable main effects with an overall consistent pattern of the results (see Online Resource 1.1). The results of the sex-stratified analysis also demonstrated a similar pattern (see Online Resource 1.2).

3.4 Machine Learning

The main analysis was performed using 2314 values per feature obtained from 1433 children. Averaged over all repetitions and folds, the model including voice features and covariates (main model) showed an *MAE* of

1.869, an R^2 of 0.120 and an r of 0.356, revealing a moderate relationship between predicted and true SDQ HI scores. Permutation testing (1000 iterations) confirmed significance (*MAE (true)* = 1.864, *MAE* (*random*) = 2.016 $p < 0.001$). Comparing the *MAE*'s from the main model to a model using only covariates (age, sex, SES, BMI-SDS) revealed superior performance of the model including voice features ($t = 1.868$; $p = 0.032$). Permutation testing (1000 iterations) for the model that used only voice features was significant (*MAE* (true) = 1.982, *MAE* (random) = 2.015, $p < 0.001$), confirming that the model performed better than chance.

Feature importance was assessed using SHAP, which estimates how much each feature contributes to the model's predictions. For the main analysis, age, SES and sex are the main predictors for SDQ HI (**Figure 3**). Several voice features contributed meaningfully to the predictions. Especially the intensity values of the *voice reset test*, i.e. *spl_quiet_V,* and the *quiet_I* condition seemed to be important voice features, aligning with the results of the LMM.

The results of the SHAP dot plot (**Figure 4**), showed that age and SES have the broadest spread, suggesting relatively strong influences on the predicted outcome. Having a lower age or SES, and being male, increased the predicted SDQ HI scores. Speaking at a higher intensity during the quiet conditions also increased the predicted SDQ HI scores. Jitter had a small average effect on prediction but showed a relatively large horizontal spread. This suggests that high Jitter values led to higher predicted SDQ HI scores.

3.4.1 ML: Impact of pubertal status

A robustness analysis was performed using 1903 values per feature obtained from 1248 children. Results confirmed that the models' predictions were reliable even when pubertal status was included as additional covariate (*MAE* = 1.902; R^2 = 0.124; r = 0.361). The result of the permutation test was significant ($p < 0.001$). SHAP values verified that demographic variables (i.e. SES, sex, pubertal status, age) were still the most important predictors of SDQ HI. The analysis reaffirmed the voice intensity during the quiet conditions, i.e. *spl_quiet_I/_V*, as a moderately important predictor.

3.4.2 ML: Sex-stratified analysis

Averaged *MAE* values from the main analysis were calculated and separated by sex (female: *MAE* = 1.749; male: *MAE* = 1.974). A *Mann−Whitney-U-Test* revealed significant differences between boys and girls (U = 731192.0; $p < 0.001$) indicating better performance in girls. However, after z-standardizing the SDQ HI scores within each sex, model performance was virtually identical for both groups (female: *MAE* = 0.751; male: *MAE* = 0.778) indicating that the observed differences are attributable to distributional differences in SDQ HI, not true sex-specific model effects.

Using two separate models the sex-stratified analysis of girls was performed using 1109 values per feature obtained from 707 girls and 1205 values per feature from 726 boys. Sex-stratified analyses showed again lower averaged *MAE* for girls (*MAE* = 1.746, R^2 = 0.101, r = 0.337) compared to boys (*MAE* = 1.985, R^2 = 0.083, r = 0.306), with both models reaching statistical significance in permutation testing ($p < 0.001$).

Averaged absolute SHAP values indicated apparent sex-specific differences in predictive patterns, with different voice features contributing to the prediction of SDQ HI in boys and girls (**Figure 5**). The model for boys indicated that besides voice intensity (condition II, V) f0 in the quiet I condition appeared important in predicting male SDQ HI. For the girls, voice intensity in the quiet conditions (I, V) seemed to be the most important voice feature.

To further examine potential sex-specific predictive effects of voice features, we included sex-by-voice feature interaction terms in the main model. Although several sex-by-voice feature interactions were selected during model fitting, residual analyses showed no statistically significant differences in model performance between boys and girls. These results indicate that, despite apparent sex differences in feature importance, the predictive value of voice features for SDQ HI is not sex-specific when controlling for general differences in voice characteristics between boys and girls. Details are provided in Online Resource 2.

4 Discussion

To our knowledge, this is the first study to analyse the potential of voice markers in children and adolescents for diagnostic purposes using machine learning in a large epidemiological sample. In LMM, we identified voice pitch (f0) and intensity, to be positively associated with SDQ HI values over different voice tasks. Using a ML approach, we found that voice features, along with age, sex, and SES, moderately predicted SDQ HI scores. Our sex-stratified analyses showed that robust prediction can be achieved with mixed-sex samples, as model performance was virtually identical for boys and girls after adjusting for variance. These findings support the use of mixed-sex samples enabling larger and more representative data collection. Together, our results provide a sound basis for further investigation of voice features as potential objective markers in the diagnostic process of ADHD in youth.

LMMs showed that children with higher SDQ HI scores exhibited higher voice intensity across all conditions and higher f0 values in conditions 2–5. This positive association indicated that children who showed more hyperactive/inattentive behaviour also display altered voice characteristics in the form of a louder/higher voice. Of note, the strongest effects were found in the *voice reset test*. This suggests that children with more pronounced symptoms of hyperactivity/inattention might find it more difficult to speak quietly once they have spoken normally/loudly. Our findings are consistent with previous studies suggesting that children with symptoms of ADHD tend to show hyperfunctional vocal behaviour, e.g. speaking louder than controls [17, 18, 23]. In contrast to the results of Hamdan et al. [23] and Garcia-Real et al. [18], we found that f0 tended to be higher, not lower, in children with higher SDQ HI scores. The positive correlation between f0 and intensity is consistent with their physiological linkage, as both are influenced by vocal effort or subglottal pressure [39, 40]. Chronic vocal misuse, for example permanently speaking too loud or with increased pitch could lead to dysphonia [41]. Within this phenomenon, vocal fold nodules (VFN) are the most common diagnosis [42]. Previous research showed that hyperactive behaviour might be an important risk factor for the development of VFN [43]. These are consequences of prolonged mechanical strain on the vocal fold epithelium due to increased tension and heightened subglottic pressure.

No significant associations were found between Jitter, MPT, or DSI and SDQ HI. Based on previous findings, Jitter values seem to be higher in children with ADHD or dysphonia [18, 44]. In addition, MPT appears to be

shorter in children with dysphonia in general or with VFN specifically [45]. There is currently no study linking hyperfunctional/inattentive behaviour to the DSI, although this index is used to assist in the diagnosis of dysphonia [46]. So far, only a few comprehensive normative values for the Dysphonia Severity Index (DSI) exist for children and adolescents. Children tend to have lower DSI values than adults - for example, 7- to 9-year-olds in Shanghai showed lower values than 18- to 23-year-olds [47]. Among Indian children, girls had significantly higher DSI values (mean 3.8 vs. 2.9), whereas no sex differences were found in the Shanghai cohort [48]. The DSI values we measured fall below this range, thus indicating a level that may suggest mild pathological changes in the voice.

Our ML model moderately predicted SDQ HI scores using voice features and relevant covariates. Voice features alone yielded significant predictive power and further enhanced model performance when combined with covariates, suggesting a biologically plausible, independent contribution to the prediction of hyperactivity/inattention symptoms. This aligns with previous studies describing the potential of voice features as diagnostic markers, including psychiatric conditions [11, 49].

SHAP analyses highlighted that voice intensity in quiet conditions contributed meaningfully to the prediction, again suggesting difficulty in modulating volume with more prominent symptoms [17, 18, 23]. While the Jitter value appears to have minimal influence on the prediction, an examination of the SHAP dot plot revealed a relatively large horizontal distribution of this feature. This indicated that high Jitter values resulted in increased predicted HI values, which is in line with findings of previous studies [18, 44]. Jitter is an acoustic index for variations in the fundamental frequency, measuring micro-variations in the periodic sound signal [46]. Among other parameters, Jitter can be used to detect the presence of a voice disorder, where higher values indicate greater variability or more irregularity in vocal fold vibration [46]. A study with 237 Egyptian children aged 4 to 12 years aimed to establish a prototype database of normal acoustic parameters. The mean jitter percentage was reported as 1.9% in boys and 1.6% in girls, with no significant sex differences observed. The study also noted a decline in jitter percentage with increasing age [50]. A prior study suggested alterations of Jitter in children with ADHD, possibly due to lower control stability of vocal folds, resulting in a less periodic larynx vibration [18].

Although initial sex-stratified analyses suggested better performance of the ML model in girls, this was attributable to lower variance in SDQ HI scores rather than true model differences. After adjusting for variance, predictive accuracy was nearly identical across sexes. This is clinically encouraging, as girls—despite being underdiagnosed—were not disadvantaged by voice-based models. While SHAP analyses in both sex-stratified and interaction models revealed sex-specific feature contributions, residual analyses showed no sex differences in model performance. This suggests that observed SHAP differences reflect normative sex-based variation in voice characteristics, rather than distinct predictive mechanisms requiring sex-specific models. Sex-specific models may therefore be unnecessary, though sex remains an important lens for interpretation and validation. In summary, sex-stratified analyses suggests that moderate prediction can be achieved using mixed-sex samples without disadvantaging either sex. In turn, this supports the use of large, diverse samples to enhance generalizability and statistical power in future studies.

In line with earlier research [36, 38], our findings confirmed that key demographic variables such as age, sex, and SES are associated with both SDQ HI scores and voice features. Including these covariates in our models

significantly improved predictive performance, underscoring the importance of accounting for known confounders when using voice-based prediction. This supports van Dellen's [4] claim that predictions should not rely solely on outcome markers, but must consider static, dynamic and contextual factors (e.g. familial, social and cultural), that shape the clinical picture. Similarly, Tan and Benos [51] emphasise the need for ML diagnostics to incorporate information that shapes a patient's health outcome, including demographic factors, to build accurate models that reflect the complexities of the real world.

4.1. Strengths and limitations

Strengths of the study include a highly standardised voice acquisition, thereby ensuring good data quality. Limitations pertain to the limited set of voice features that were included in this analysis. While serial speech (reciting number words) is a well-standardized task for children and adolescents, it does not reflect the characteristics of everyday communication. Therefore, additional investigations using standardized texts (e.g., the Rainbow Passage) and recordings of spontaneous speech are warranted in future studies.

5 Conclusion

Within the framework of precision psychiatry, the results of this study provide first insights into the identification and potential use of voice markers in children and adolescents. Our results indicate the added value of voice features to the mostly subjective evaluation of symptoms of hyperactivity/inattention and stress the need of further investigating the potential of voice-based ML models to detect childhood psychiatric disorders. Research on adults has already confirmed voice features as potential diagnostic markers in psychiatric populations [52]. In adult ADHD, previous research has shown that vocal features, including f0 and jitter, can be used to classify subtypes [53]. The prospect that paediatric voice features could also be used as markers to support a more accurate diagnostic process remains to be explored but seems promising. In summary, we identified voice markers with predictive value in the general population, warranting further investigation in clinical samples.

Declarations

Contributions

Corresponding author

Correspondence to Rachel Bamberger.

Ethics declaration

Ethical approval

The study was approved by the ethics board of the *University of Leipzig* (Reg. No. 477/19-ek) and adheres to the Declaration of Helsinki

Informed consent

Written informed consent was received from all parents and children aged over 12 years.

Competing interests

All authors have no conflicts of interest to declare that are relevant to the content of this article.

Data/Code availability

The data that support the findings of this study are not openly available due to reasons of sensitivity. The codes used for statistical analyses are available in the following GitHub repository: https://github.com/RaBa-fzj/VoicesOfChange.git

References

1. Dalsgaard S, Thorsteinsson E, Trabjerg BB et al (2020) Incidence Rates and Cumulative Incidences of the Full Spectrum of Diagnosed Mental Disorders in Childhood and Adolescence. JAMA Psychiatry 77:155-164. https://doi.org/10.1001/jamapsychiatry.2019.3523.
2. Merikangas KR, He JP, Burstein M et al (2010) Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication--Adolescent Supplement (NCS-A). J Am Acad Child Adolesc Psychiatry 49:980-989. https://doi.org/10.1016/j.jaac.2010.05.017.
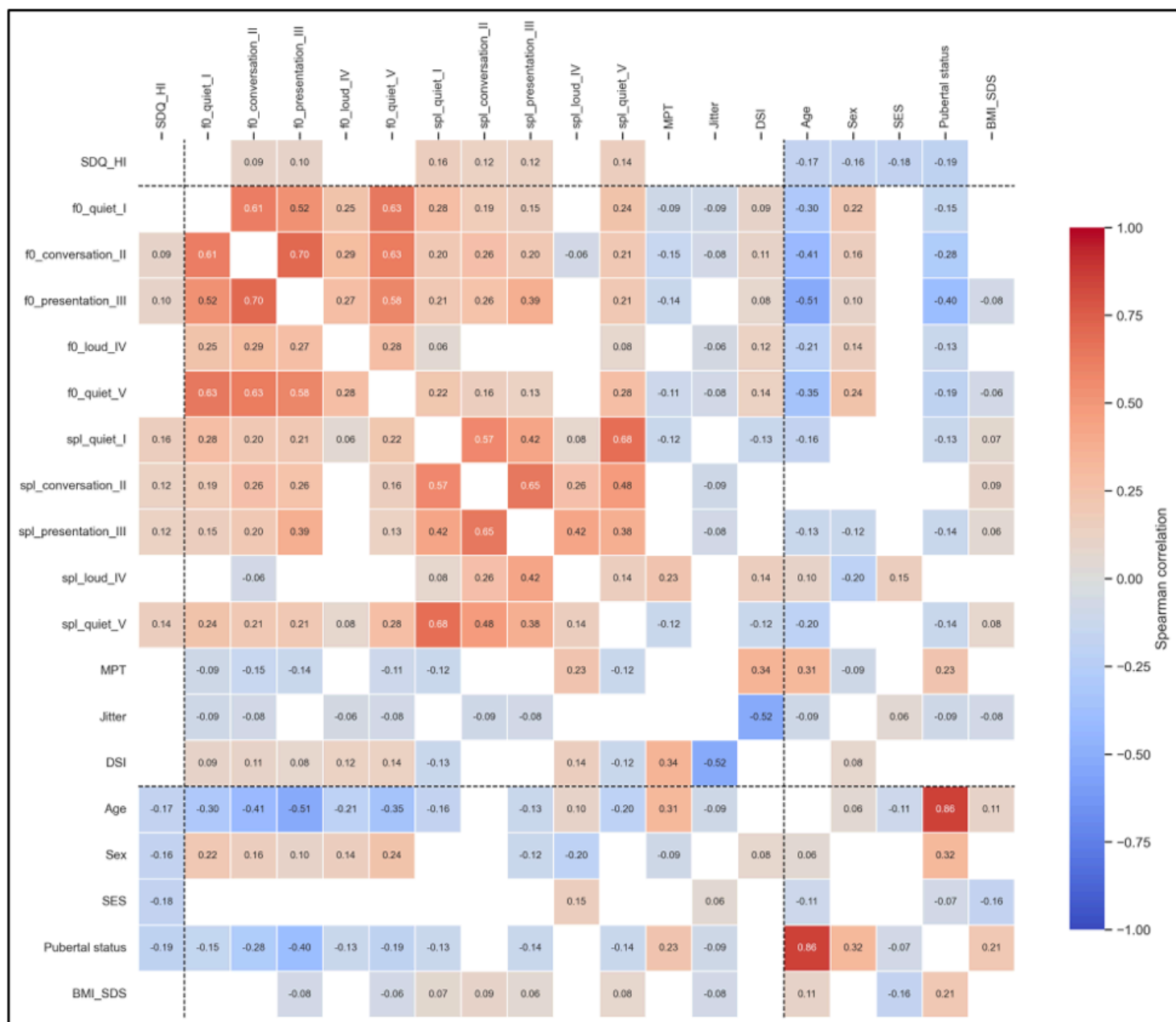
3. Gauld C, Viaux-Savelon S, Falissard B, Fourneret P (2024) Precision child and adolescent psychiatry: reductionism, fad, or change of identity of the discipline? Eur Child Adolesc Psychiatry 33:1193-1196. https://doi.org/10.1007/s00787-023-02240-6.

4. Van Dellen E (2024) Precision psychiatry: predicting predictability. Psychol Med 54:1500-1509. https://doi.org/10.1017/S0033291724000370.

5. Leikauf JE (2024) Precision Psychiatry Implementation in Child and Adolescent Psychiatric Clinical Practice: Opportunities and Challenges. Psychiatric Annals 54. https://doi.org/10.3928/00485713-20240314-01.

6. Posner J (2018) The Role of Precision Medicine in Child Psychiatry: What Can We Expect and When? J Am Acad Child Adolesc Psychiatry 57:813-817. https://doi.org/10.1016/j.jaac.2018.07.874.

7. Fuentes-Aguilar RQ, Pérez-Espinosa H, Filigrana-De-La-Cruz MA (2022) Biosignals analysis (heart, phonatory system, and muscles). In: Biosignal Processing and Classification Using Computational Learning and Intelligence. p 7-26

8. Almaghrabi SA, Clark SR, Baumert M (2023) Bio-acoustic features of depression: A review. Biomedical Signal Processing and Control 85. https://doi.org/10.1016/j.bspc.2023.105020.

9. Li S, Nair R, Naqvi SM (2024) Acoustic and Text Features Analysis for Adult ADHD Screening: A Data-Driven Approach Utilizing DIVA Interview. IEEE J Transl Eng Health Med 12:359-370. https://doi.org/10.1109/JTEHM.2024.3369764.

10. Olah J, Diederen K, Gibbs-Dean T, Kempton MJ, Dobson R, Spencer T, Cummins N (2023) Online speech assessment of the psychotic spectrum: Exploring the relationship between overlapping acoustic markers of schizotypy, depression and anxiety. Schizophr Res 259:11-19. https://doi.org/10.1016/j.schres.2023.03.044.

11. Fagherazzi G, Fischer A, Ismael M, Despotovic V (2021) Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice. Digit Biomark 5:78-88. https://doi.org/10.1159/000515346.

12. Low DM, Bentley KH, Ghosh SS (2020) Automated assessment of psychiatric disorders using speech: A systematic review. Laryngoscope Investig Otolaryngol 5:96-116. https://doi.org/10.1002/lio2.354.

13. Rogers HP, Hseu A, Kim J et al (2024) Voice as a Biomarker of Pediatric Health: A Scoping Review. Children (Basel) 11. https://doi.org/10.3390/children11060684.

14. Von Polier GG, Ahlers E, Volkening J et al (2025) Exploring voice as a digital phenotype in adults with ADHD. Sci Rep 15:18076. https://doi.org/10.1038/s41598-025-01989-x.

15. Dum R, Ghahramani A, Baweja R, Bellon A (2022) Dopamine Receptor Expression and the Pathogenesis of Attention-Deficit Hyperactivity Disorder: a Scoping Review of the Literature. Current Developmental Disorders Reports 9:127-136. https://doi.org/10.1007/s40474-022-00253-5.

16. Turk AZ, Lotfi Marchoubeh M, Fritsch I, Maguire GA, Sheikhbahaei S (2021) Dopamine, vocalization, and astrocytes. Brain Lang 219:104970. https://doi.org/10.1016/j.bandl.2021.104970.

17. Barona-Lleo L, Fernandez S (2016) Hyperfunctional Voice Disorder in Children With Attention Deficit Hyperactivity Disorder (ADHD). A Phenotypic Characteristic? J Voice 30:114-119. https://doi.org/10.1016/j.jvoice.2015.03.002.

18. Garcia-Real T, Diaz-Roman TM, Garcia-Martinez V, Vieiro-Iglesias P (2013) Clinical and acoustic vocal profile in children with attention deficit hyperactivity disorder. J Voice 27:787 e711-788. https://doi.org/10.1016/j.jvoice.2013.06.013.

19. D'alatri L, Petrelli L, Calò L, Picciotti PM, Marchese MR, Bussu F (2015) Vocal fold nodules in school age children: attention deficit hyperactivity disorder as a potential risk factor. J Voice 29:287-291. https://doi.org/10.1016/j.jvoice.2014.07.019.

20. Erdur O, Herguner A, Ozturk K, Kibar E, Elsurer C, Bozkurt MK, Herguner S (2016) Attention deficit hyperactivity disorder symptoms in children with vocal fold nodules. Int J Pediatr Otorhinolaryngol 85:5-7. https://doi.org/10.1016/j.ijporl.2016.03.017.

21. Buitelaar J, Bolte S, Brandeis D et al (2022) Toward Precision Medicine in ADHD. Front Behav Neurosci 16:900981. https://doi.org/10.3389/fnbeh.2022.900981.

22. Woerner W, Becker A, Rothenberger A (2004) Normative data and scale properties of the German parent SDQ. Eur Child Adolesc Psychiatry 13 Suppl 2:Ii3-10. https://doi.org/10.1007/s00787-004-2002-6.

23. Hamdan AL, Deeb R, Sibai A, Rameh C, Rifai H, Fayyad J (2009) Vocal characteristics in children with attention deficit hyperactivity disorder. J Voice 23:190-194. https://doi.org/10.1016/j.jvoice.2007.09.004.

24. Poulain T, Baber R, Vogel M et al (2017) The LIFE Child study: a population-based perinatal and pediatric cohort in Germany. Eur J Epidemiol 32:145-158. https://doi.org/10.1007/s10654-016-0216-9.

25. Lampert T, Müters S, Stolzenberg H, Kroll LE (2014) Messung des sozioökonomischen Status in der KiGGS-Studie. Bundesgesundheitsblatt 57:767-760. https://doi.org/10.1007/s00103-014-1974-8.

26. Kromeyer-Hauschild K, Wabitsch M, Kunze D et al (2001) Perzentile für den Body-mass-Index für das Kindes- und Jugendalter unter Heranziehung verschiedener deutscher Stichproben. Monatsschrift Kinderheilkunde 149:807-818. https://doi.org/10.1007/s001120170107.

27. Marshall WA, Tanner JM (1969) Variations in pattern of pubertal changes in girls. Arch Dis Child 44:291-303. https://doi.org/10.1136/adc.44.235.291.

28. Marshall WA, Tanner JM (1970) Variations in the pattern of pubertal changes in boys. Arch Dis Child 45:13-23. https://doi.org/10.1136/adc.45.239.13.

29. Schutte HK, Seidner W (2009) Recommendation by the Union of European Phoniatricians (UEP): Standardizing Voice Area Measurement/Phonetography. Folia Phoniatrica et Logopaedica 35:286-288. https://doi.org/10.1159/000265703.

30. Berger T, Peschel T, Vogel M et al (2019) Speaking Voice in Children and Adolescents: Normative Data and Associations with BMI, Tanner Stage, and Singing Activity. J Voice 33:580 e521-580 e530. https://doi.org/10.1016/j.jvoice.2018.01.006.

31. Poulain T, Fuchs M, Vogel M, Jurkutat A, Hiemisch A, Kiess W, Berger T (2020) Associations of Speaking-Voice Parameters With Personality and Behavior in School-Aged Children. J Voice 34:485 e423-485 e431. https://doi.org/10.1016/j.jvoice.2018.09.022.

32. Wuyts FL, De Bodt MS, Molenberghs G et al (2000) The dysphonia severity index: an objective measure of vocal quality based on a multiparameter approach. J Speech Lang Hear Res 43:796-809. https://doi.org/10.1044/jslhr.4303.796.

33. Sobol M, Sielska-Badurek EM (2022) The Dysphonia Severity Index (DSI)-Normative Values. Systematic Review and Meta-Analysis. J Voice 36:143.e149-143.e113. https://doi.org/10.1016/j.jvoice.2020.04.010.

34. Pedregosa F, Varoquaux G, Gramfort A et al (2011) Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12:2825-2830. https://doi.org/https://jmlr.csail.mit.edu/papers/volume12/pedregosa11a/pedregosa11a.pdf.

35. Nadeau C, Bengio Y (1999) Inference for the generalization error. Advances in neural information processing systems 12.

36. Mennies RJ, Birk SL, Norris LA, Olino TM (2021) The Main and Interactive Associations between Demographic Factors and Psychopathology and Treatment Utilization in Youth: A Test of Intersectionality in the ABCD Study. Research on Child and Adolescent Psychopathology 49:5-17. https://doi.org/10.1007/s10802-020-00687-8.

37. Reiss F (2013) Socioeconomic inequalities and mental health problems in children and adolescents: A systematic review. Social Science & Medicine 90:24-31. https://doi.org/https://doi.org/10.1016/j.socscimed.2013.04.026.

38. Kim KH, Kim RB, Hwang DU, Won SJ, Woo SH (2016) Prevalence of and Sociodemographic Factors Related to Voice Disorders in South Korea. Journal of Voice 30:246.e241-246.e247. https://doi.org/10.1016/j.jvoice.2015.04.010.

39. Jessen M, Koster O, Gfroerer S (2005) Influence of vocal effort on average and variability of fundamental frequency. The International Journal of Speech, Language and the Law 12:174-213. https://doi.org/10.1558/sll.2005.12.2.174.

40. Tilsen S (2016) A shared control parameter for F0 and intensity. Proceedings of Speech Prosody 2016:1066-1070. https://doi.org/10.21437/SpeechProsody.2016-219.

41. Martins RHG, Hidalgo Ribeiro CB, Fernandes De Mello BMZ, Branco A, Tavares ELM (2012) Dysphonia in Children. Journal of Voice 26:674.e617-674.e620. https://doi.org/10.1016/j.jvoice.2012.03.004.

42. Milea A, Milea A, Sarafoleanu C (2024) Dysphonia in children. Romanian Journal of Rhinology 14:166-171. https://doi.org/10.2478/rjr-2024-0025.

43. Reis-Rego Â, Santos PH, Santos G et al (2019) Behavioral Profile of Children With Vocal Fold Nodules-A Case-control Study. J Voice 33:584.e581-584.e584. https://doi.org/10.1016/j.jvoice.2018.02.009.

44. Meredith ML, Theis SM, Mcmurray JS, Zhang Y, Jiang JJ (2008) Describing pediatric dysphonia with nonlinear dynamic parameters. Int J Pediatr Otorhinolaryngol 72:1829-1836. https://doi.org/10.1016/j.ijporl.2008.09.004.

45. Gramuglia AC, Tavares EL, Rodrigues SA, Martins RH (2014) Perceptual and acoustic parameters of vocal nodules in children. Int J Pediatr Otorhinolaryngol 78:312-316. https://doi.org/10.1016/j.ijporl.2013.11.032.

46. Pädaudiologie DGFRPU (2022) Diagnostik und Therapie von Störungen der Stimmfunktion (Dysphonien). AWMF. https://register.awmf.org/assets/guidelines/049-008l_S2k_Diagnostik-Therapie-Stoerungen-der-Stimmfunktion-Dysphonien_2023-01.pdf.

47. Kim H, Gao S, Shi R, Zhang Y, Liu X, Yi B (2019) Influence of gender and age on the Dysphonia Severity Index: A normative study in a Shanghainese population. Clin Linguist Phon 33:279-293. https://doi.org/10.1080/02699206.2018.1508309.

48. Pebbili GK, Kidwai J, Shabnam S (2017) Dysphonia Severity Index in Typically Developing Indian Children. J Voice 31:125.e121-125.e126. https://doi.org/10.1016/j.jvoice.2015.12.017.

49. Sara JDS, Orbelo D, Maor E, Lerman LO, Lerman A (2023) Guess What We Can Hear-Novel Voice Biomarkers for the Remote Detection of Disease. Mayo Clin Proc 98:1353-1375. https://doi.org/10.1016/j.mayocp.2023.03.007.

50. Ibrahim RA, Hassan EM (2021) Normative vocal acoustic parameters for preschool and elementary school Egyptian children. Int J Pediatr Otorhinolaryngol 143:110662. https://doi.org/10.1016/j.ijporl.2021.110662.

51. Tan MJT, Benos PV (2025) Addressing Intersectionality, Explainability, and Ethics in AI-Driven Diagnostics: A Rebuttal and Call for Transdiciplinary Action. arXiv preprint arXiv:2501.08497. https://doi.org/10.48550/arXiv.2501.08497.

52. Menne F, Dörr F, Schräder J, Tröger J, Habel U, König A, Wagels L (2024) The voice of depression: speech features as biomarkers for major depressive disorder. BMC Psychiatry 24:794. https://doi.org/10.1186/s12888-024-06253-6.

53. Bruno E, Martz E, Weiner L, Greco A, Vanello N (2023) Speech signal analysis as an aid to clinical diagnosis and assessment of mental health disorders. Biomedical Signal Processing and Control 85:104854. https://doi.org/https://doi.org/10.1016/j.bspc.2023.104854.
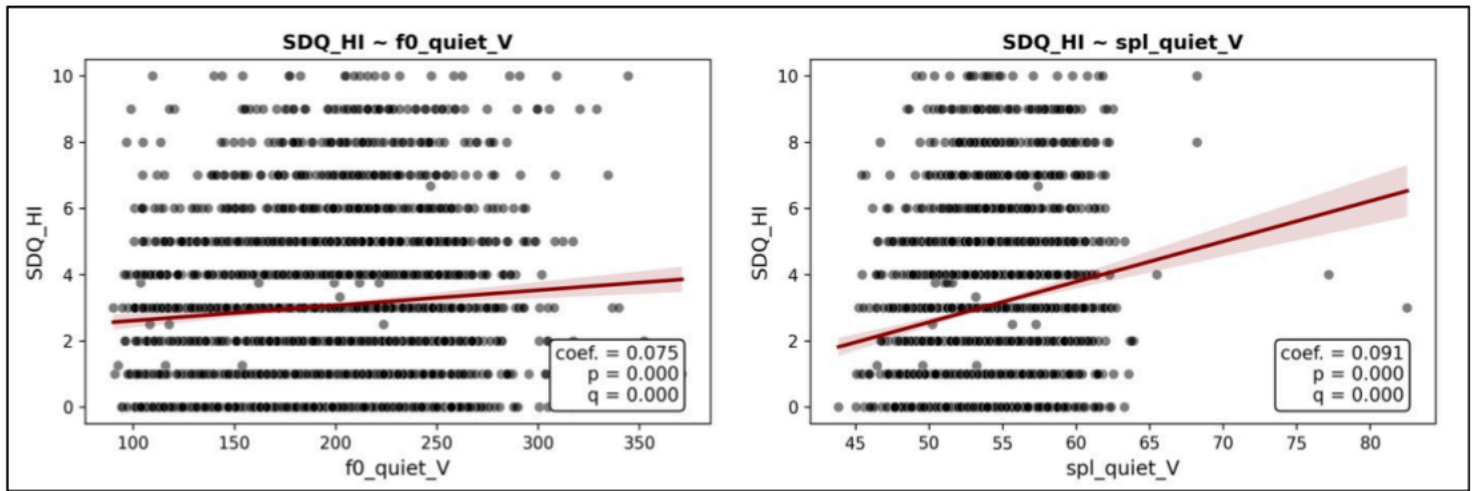
Figures

Figure 1

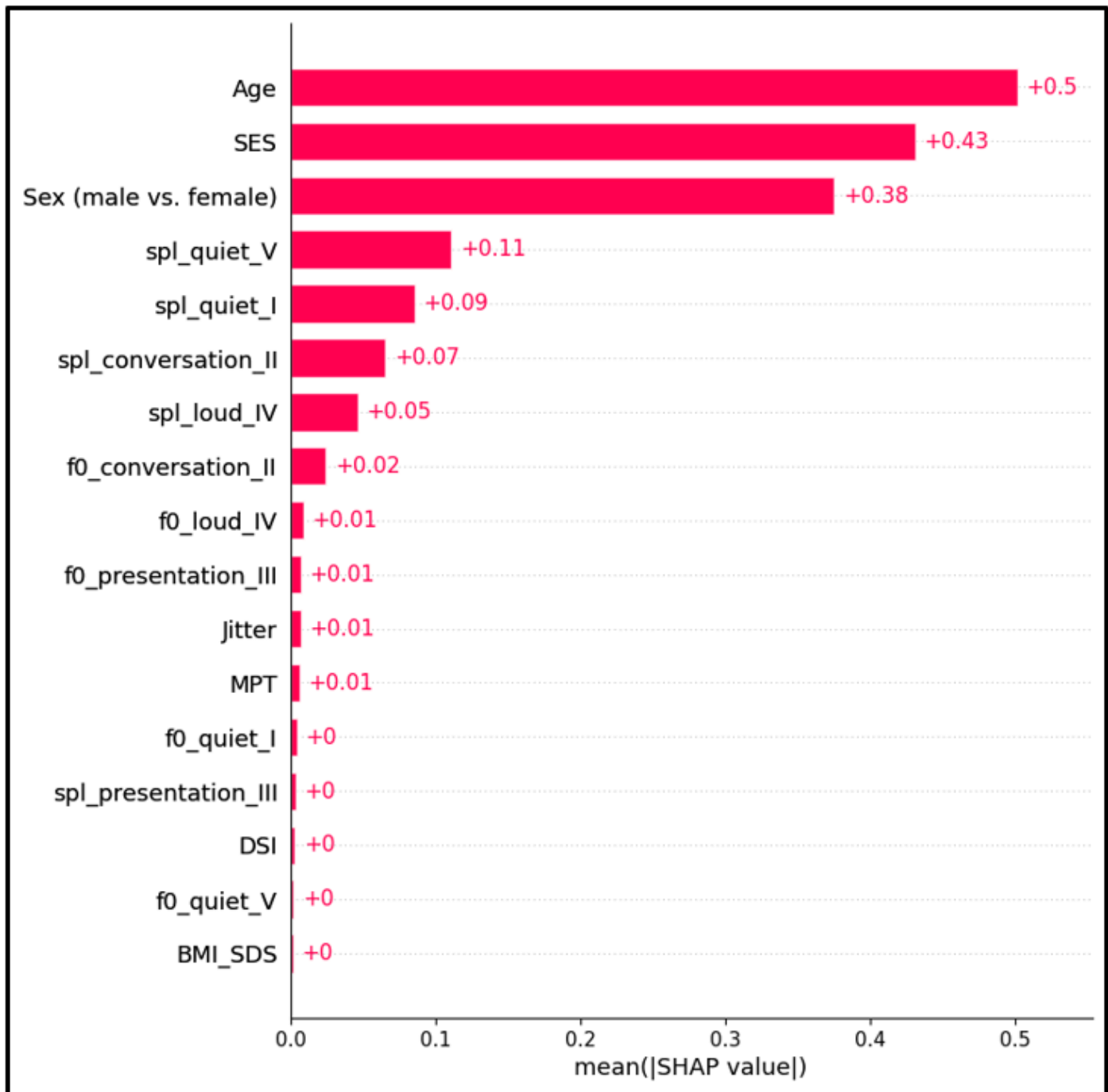Correlation matrix of significant correlations after FDR-correction (Spearman's Rho)

Note. Only statistically significant correlations (FDR-corrected $p < 0.05$) are displayed in the figure; SDQ HI = Strengths and Difficulties Questionnaire subscale hyperactivity/inattention; f0 = fundamental frequency; spl = sound pressure level (intensity); MPT = maximum phonation time; DSI = dysphonia-severity-index; SES = socio-economic status; BMI_SDS = BMI-standard deviation score
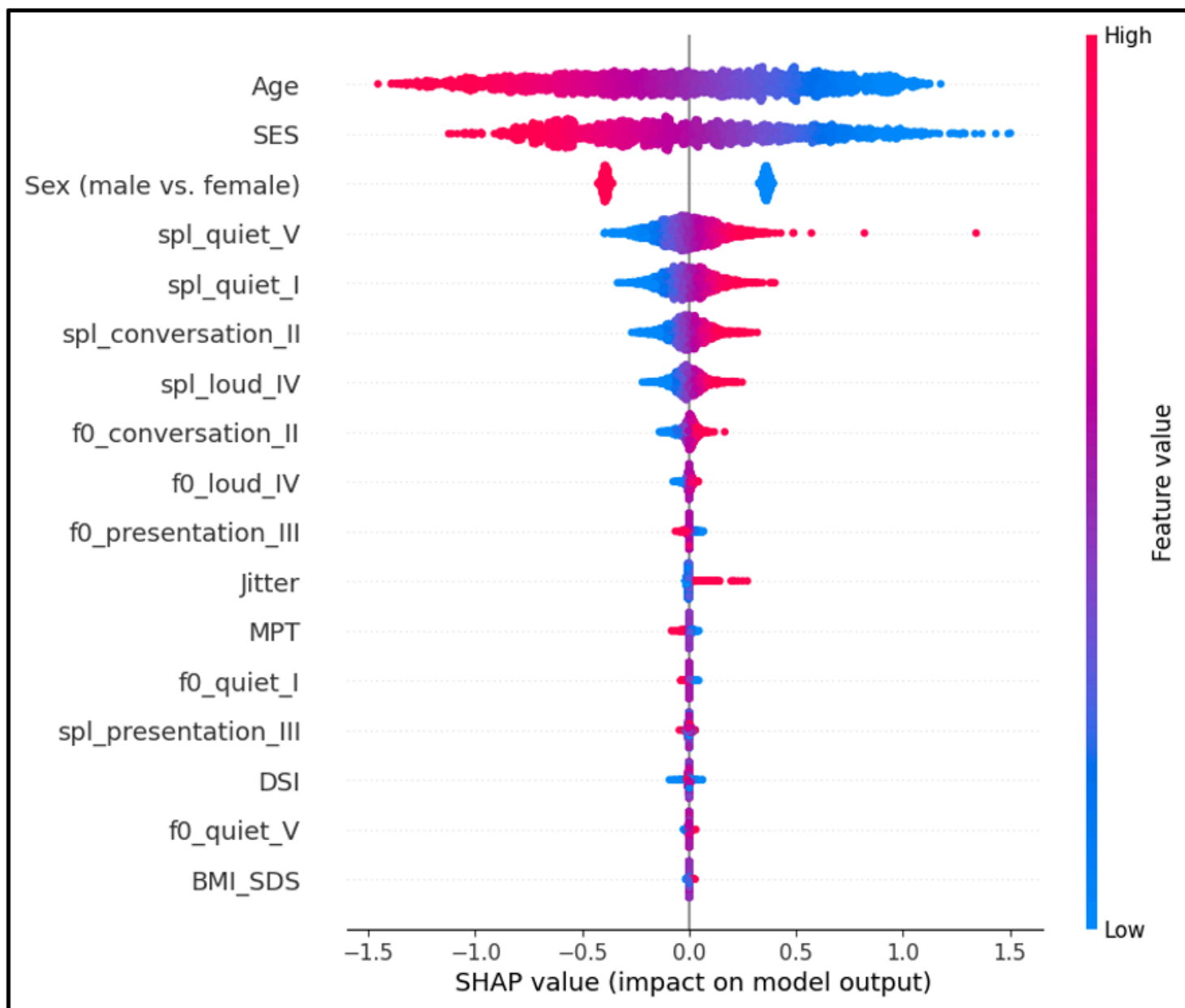
Figure 2

Scatterplot SDQ HI and f0/spl_quiet_V

*Note.*SDQ HI = Strengths and Difficulties Questionnaire subscale hyperactivity/inattention; f0 = fundamental frequency; spl = sound pressure level (intensity); coef = coefficient; p/q = p-/q-value (FDR adjusted p-value)

Figure 3

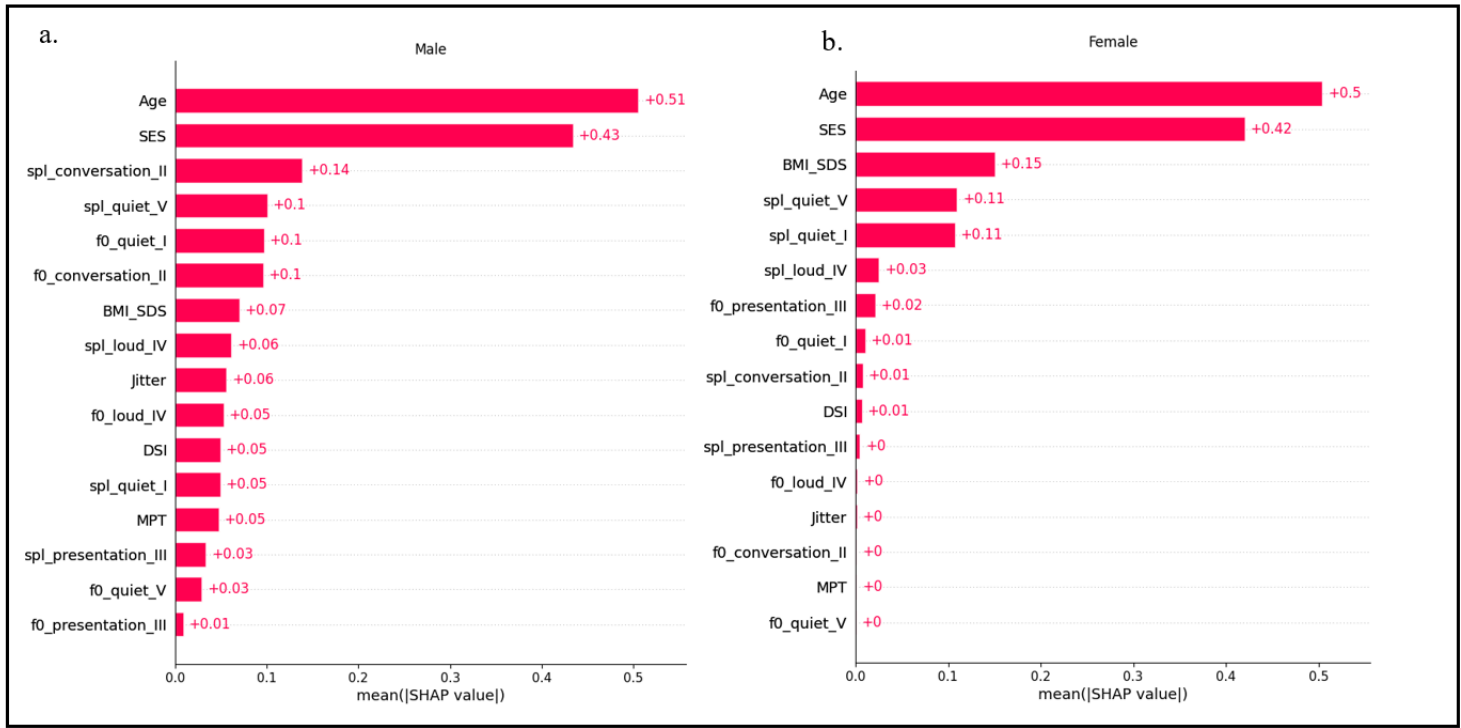SHAP bar plot main analysis (Ridge Regression)

*Note.*SHAP = Shapley Additive Explanation; SES = socio-economic status

Figure 4

SHAP dot plot main analyis (Ridge Regression)

Note. SHAP = Shapley Additive Explanation; SES = socio-economic status; MPT = Maximum phonation time; DSI = Dysphonia-severity-index; BMI_SDS = BMI-standard deviation score

Figure 5

SHAP bar plot sex-stratified analysis (Ridge Regression)

Note. a = Boys, b = Girls; SHAP = Shapley Additive Explanation; SES = socio-economic status

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- OnlineResource1.xlsx
- OnlineResource2.pdf