

Speech-Based Markers in Paediatric ADHD: A Longitudinal Case-Control Study of Voice Features and Medication Effects

Rachel Bamberger^{1,2,3} (0009-0008-0136-6553)

Gianna Kuhles^{1,4} (0000-0002-1192-4888)

Leon D. Lotter^{1,4,5} (0000-0002-2337-6073)

Juergen Dukart^{1,4} (0000-0003-0492-5644)

Kerstin Konrad^{3,6} (0000-0001-9039-2615)

Thomas Günther⁷ (0000-0002-9369-2450)

Michael Siniatchkin²

Michael Fuchs⁸ (0000-0002-9491-7295)

Georg von Polier^{1,9,2} (0000-0002-2739-8528)

1 Institute of Neuroscience and Medicine, Brain & Behaviour (INM-7), Research Centre Jülich, Jülich, Germany

2 Department of Child and Adolescent Psychiatry, Psychotherapy and Psychosomatics, University Hospital RWTH Aachen, Aachen, Germany

3 Institute of Medical Psychology and Medical Sociology, Faculty of Medicine, RWTH Aachen University, Aachen, Germany

4 Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University, Düsseldorf, Germany

5 Max Planck School of Cognition, Stephanstrasse 1A, 04103 Leipzig, Germany

6 Institute of Neuroscience and Medicine, JARA-Institute Brain structure-function relationships (INM-10), Research Centre Jülich, Jülich, Germany

7 Department of Psychological Assessment and Intervention, Institute of Psychology, RWTH Aachen University, Aachen, Germany

8 Section of Phoniatics and Audiology, Department of Otorhinolaryngology, University Hospital Leipzig, Leipzig, Germany

9 Department of Child and Adolescent Psychiatry, Psychotherapy and Psychosomatics, University Hospital Leipzig, Leipzig, Germany

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Abstract

Background

Diagnosis and treatment monitoring of attention-deficit/hyperactivity disorder (ADHD) largely rely on subjective assessments, highlighting the need for objective markers. Voice features and speech embeddings represent promising candidates for such markers, as they may capture alterations in speech production relevant to ADHD. However, it remains unclear which speech features are most informative for distinguishing ADHD and monitoring treatment effects, and which speech tasks most reliably elicit such differences.

Methods

Twenty-seven children with ADHD and 27 age-matched neurotypical controls completed six speech tasks across two study visits. Children with ADHD were unmedicated at baseline (first visit) and were assessed under prescribed methylphenidate treatment at follow-up, whereas controls underwent repeated assessment without intervention. Established acoustic voice features (eGeMAPS) and high-dimensional speech embeddings (WavLm, Whisper) were extracted and analysed using linear mixed models to examine baseline group differences and group-by-time interaction effects reflecting medication-associated change patterns.

Results

At baseline, children with ADHD differed significantly from controls in frequency, spectral, and temporal voice features, characterized by lower and more variable pitch, altered spectral properties, and reduced rhythmic stability. Group-by-time interaction effects indicated medication-associated modulation in the ADHD group, including reduced loudness variability and increased precision of vowel articulation at follow-up, changes not observed in controls. Speech embeddings revealed additional baseline and interaction effects beyond established acoustic features. Free speech tasks, particularly picture description, yielded the most robust and consistent effects.

Conclusion

Children with ADHD differed from neurotypical controls in vocal features at baseline and showed distinct longitudinal change patterns consistent with medication-related change. These findings support further investigation of speech-based measures as candidate digital phenotypes and potential digital biomarkers in ADHD, with picture description emerging as a particularly promising task for future clinical assessment protocols.

Keywords: ADHD; Voice; Speech; Speech tasks; Children; Methylphenidate; Digital biomarkers; Digital phenotyping

1 Introduction

Diagnosis and treatment evaluation of attention-deficit/hyperactivity disorder (ADHD) rely largely on subjective assessments by clinicians, parents, and patients [1]. Limitations thereof include high observer bias as well as insufficient validity or reliability [2]. This has led to a broad consensus on the need for more objective markers [1, 3, 4]. Objective markers may provide additional information for diagnosis and facilitate more personalised ADHD treatment [2, 5].

Speech production is one of the most complex and elaborate motor outputs of the brain [6]. The voice conveys information beyond plain content and could therefore provide non-invasive insights into neurodevelopmental disorders [7]. Research into voice features as biomarkers within psychiatry has steadily increased, and their use appears promising [8, 9].

Deficits in neuropsychological mechanisms, including motor control or cognitive processes, could alter speech patterns. For example, the necessary synchronisation between motor output and sensory perception required for speech production seems to be different in individuals with ADHD, leading to a faster speaking rate and alternating motion ranges [10]. Furthermore, deficits in inhibitory control in ADHD seem to impact the speech production system requiring this control mechanism to prevent or correct errors while speaking [11]. Research on voice in children with ADHD shows more hyperfunctional vocal behaviour, i.e. speaking with a louder voice or applying more pressure [12] as well as more hoarseness, straining and breathiness [13, 14]. Our previous research in clinical and population samples has demonstrated an association between voice features, such as pitch and intensity, and ADHD symptoms in children and adults [15, 16]. Preliminary evidence suggests that voice features may be sensitive to pharmacological modulation with methylphenidate (MPH) supporting their potential utility for treatment monitoring. Specifically, changes in fundamental frequency (F0) [17] and jitter (an acoustic parameter of F0 stability) [18] were observed following MPH administration.

Established voice features allow hypothesis-driven investigation of specific acoustic parameters, but do not capture other speech features known to be compromised in ADHD. Speech embeddings provide an additional, exploratory approach by encoding high-dimensional speech representations that integrate multiple aspects of speech, including (para)linguistic cues, prosody, and temporal dynamics [19, 20]. In this sense, embeddings can be used to explore whether speech contains additional relevant information beyond established interpretable voice features.

Importantly, both voice features and speech-embeddings depend critically on the speech context in which they are elicited. In clinical practice, multiple speech tasks are used to capture distinct vocal and speech information. Free speech paradigms, such as picture description, are the most commonly applied and highly standardised [21]. Other frequently used tasks include sustained phonation, diadochokinesia, and reading.

The present study aimed to evaluate the potential utility of speech-based measures in a clinical sample by (a) comparing voice features and speech embeddings between unmedicated children with ADHD and age-matched neurotypical controls and (b) examining longitudinal changes in these measures following MPH treatment in children with ADHD, compared with repeated assessments without treatment in controls. It is important to note that this study primarily focused on established voice features. As speech embeddings contain information that is not directly interpretable, analysis remained largely exploratory. Additionally, we systematically investigated which speech tasks were most sensitive to differences and changes in speech features, to inform future task selection. Based on the existing literature, we hypothesised that speech features contain information relevant to distinguishing between children with ADHD and neurotypical controls, as well as to detecting MPH treatment response. We also hypothesised that certain speech tasks are better at revealing potential differences or changes in speech than others. Due to the heterogeneity of previous findings, no directional hypotheses were formulated.

2 Methods

2.1 Study design and participants

We conducted a case-control study with a pre-post design. The data were acquired at the Department of *Child and Adolescent Psychiatry, Psychotherapy and Psychosomatics at RWTH Aachen University Hospital*. The study is in accordance with the Declaration of Helsinki, was approved by the *Ethics Committee of the Medical Faculty of RWTH Aachen University* (approval no. 032/20), and registered with the *German Clinical Trials Registry* (DRKS; registration no. DRKS00030766). Written informed consent was received from all parents and children, regardless of age.

We included children aged between 8 and 16 years who were either neurotypical or had a clinical ADHD diagnosis according to DSM-5 criteria and had been prescribed pharmacotherapy with MPH. Children with an IQ below 80 or a comorbid autism spectrum disorder were excluded. For the present study, boys who already underwent pubertal voice change (fundamental frequency below C3, i.e., 130.82 Hz [22]) were excluded from the analysis. Children with ADHD were recruited from within the clinic and local outpatient practices, neurotypical controls (NC) through public advertisements (flyers/mailling lists).

Participation involved two visits (baseline/follow-up) at our clinic, ideally scheduled about 8 weeks apart. These included an IQ estimation (if a full assessment was not available), neuropsychological and speech tasks, questionnaires and, for eligible and willing participants, an optional MRI session. For the present study, data from the speech tasks (\approx 20 minutes) and questionnaires were used, as described further below.

A total of 57 children participated in the study. Three males of the control group had already undergone voice change, so their data were excluded. The remaining 54 children (female = 31; male = 21, diverse = 2; ADHD = 27; NC = 27) had a mean age of 12.40 years (range: 8.08 - 16.92). Out of the whole sample, only two children with ADHD discontinued pharmacotherapy and were therefore not assessed at follow-up.

2.2 Measurements

2.2.1 Demographic and health data

Demographic (age, gender) and health data were collected through interviews and questionnaires, diagnostic information (ADHD subtype and comorbidities according to ICD-10 criteria) was obtained from clinic records. To assess whether DSM-5 criteria for ADHD were met, parents completed the *Attention Deficit Hyperactivity Disorder* supplement of the *Schedule for Affective Disorders and Schizophrenia for School-Age Children—*

Present and Lifetime Version (K-SADS-PL) [23]. Absence of a diagnoses in controls was confirmed by the parents and validated by evaluation of the K-SADS-PL.

At both visits, parents completed the long form of the German adaption of the *Conners 3rd Edition-Parent* (Conners 3-P) [24]. This questionnaire is designed to assess symptoms of ADHD and their severity in children aged 6–18. For the present study, we used the values of children with ADHD on the *Inattention* (IA) and *Hyperactivity/Impulsivity* (HI) scales.

2.2.2 Voice measurements

Voice measurements were conducted according to a standard operating procedure (SOP), which is guideline recommended [25, 26]. The SOP included detailed written instructions regarding equipment and setup, recording parameters, and verbal instructions for each task to ensure reproducible voice recordings (**Appendix S1**).

During data collection, the recording setup was updated in line with current guideline recommendations (**Appendix S2**). Consequently, ten participants (9 with ADHD) were recorded using the earlier setup. The initial setup employed a cardioid microphone (*Elgato Wave:1*) with a microphone-mouth distance between 30-40 cm. The final recording setup used an omnidirectional headset microphone (*Sennheiser HSP 2 EW 3*) positioned 2 cm from the corner of the mouth. All recordings were obtained at a sample rate of 48.000 Hz at 24-bit resolution and were stored in the .wav audio format [25].

Recordings were conducted in a quiet room with the child and the researcher seated opposite each other. During both visits, children were asked to complete a total of six speech tasks in German (**Table 1**). The speech task protocol was slightly refined in the early stages of the data collection to reflect updated methodological recommendations. A five-second silent recording was added, to analyse ambient noise levels during each visit [26]. Two sustained phonation tasks (normal/loud voice) were included to enable reliable analysis of jitter values [27]. Additional recordings were not taken for the first 5 children (ADHD) who participated.

2.2.3 Voice features

Voice features were automatically extracted from the speech recordings using the standardised and widely used *eGeMAPS* feature set [28], which is implemented in the *openSMILE* toolkit [29]. The *eGeMAPS* set comprises 88 theoretically grounded voice features, extracted separately from each task. These features are derived from four voice-domains: Frequency (e.g. pitch), Energy/amplitude (e.g. loudness), Spectral (e.g. Mel-Frequency Cepstral Coefficients), and Temporal (e.g. mean length of voiced regions).

2.2.4 Speech embeddings

Speech embeddings were automatically extracted using the *WavLM* [30] and *Whisper* [31] base encoders, resulting in 768-dimensional representations per child, task, and visit. These transformer-based models generate high-dimensional representations of speech that capture e.g. (para)linguistic, prosodic, and temporal characteristics. Both models are trained on large, varied datasets, so their embeddings reflect multidimensional speech dynamics and may reveal subtle speech changes relevant for clinical or behavioural characterisation [19].

2.3 Statistical Analysis

2.3.1 Data preparation and quality control

The audio recordings were preprocessed using *Python* (3.9) and custom scripts that leveraged libraries including *librosa* (0.11) [32], *praat-parselmouth* (0.4) [33], and *webrtcvad* (2.0). All files were resampled to 48.000 kHz in a 16-bit format to comply with the preprocessing packages and feature extraction methods. The recordings were cut according to the respective task (**Appendix S3**) and the targeted recording length (**Table 1**). Each recording's audio wave was visually inspected and a minimum of 10% of recordings from each task was fully listened to and checked for inconsistencies. Quality control of ambient noise levels and signal-to-noise-ratio (SNR) was performed using the recordings of silence (ambient noise) and phonation normal (SNR). All recordings met the guideline described quality criteria [26, 34] (**Appendix S4**). Following preprocessing and quality control, audio files were passed to the feature extraction pipelines using Python scripts as described above.

2.3.2 Linear mixed models

We used linear mixed models (LMMs) to assess effects of group at baseline as well as group-by-time interactions on voice features. All variables were z-standardized independently across subjects and sessions. Intensity-based features (i.e., loudness-related) were set to missing (*NaN*) for recordings without the headset microphone as these features are sensitive to recording distance and equipment. Separate LMMs were fitted for each of the 88 voice features extracted from each task. Features exhibiting zero variance were excluded from analysis, resulting in a total of 512 models. Models were specified in a 2x2 design with group (controls/ADHD; reference: controls) and time (baseline/follow-up; reference: baseline) as main effects, their interaction, as well as age at baseline and gender (m/w/d) as covariates [35]. Using this design, the main effect of group can be interpreted as the difference between ADHD and controls at baseline. We interpreted significant interaction effect as indication for a difference in changes observed over time. Random intercepts accounted for within-subject variability. Features were excluded if they were structurally invalid for the task (i.e. unvoiced features in phonation tasks) or showed zero

variance. LMMs were fitted using the python *statsmodels* (0.14.5) package and model fit was optimised using a standard procedure (*Powell* method, 1000 iterations). Multiple testing was corrected task-wise using the False Discovery Rate (FDR; Benjamini-Hochberg procedure). Results with FDR-corrected p-values (q-values) < 0.05 were considered statistically significant. To assess robustness to outliers that may occur due to the high inter- and intrapersonal variability of voice [36], the LMM analyses were repeated using ranked data, resembling a rank-based Wilcoxon-like procedure [37]. As LMM estimates do not typically allow for standardised effect size measures, the fixed-effects regression coefficient (β) was used as a measure of effect magnitude. Previously done z-transformation ensured that the coefficients were comparable among predictors. Additionally, we reran the analyses excluding three children with ADHD assessed during their medication pause and robust interaction effects were re-tested by including the *time between sessions* as an additional covariate. Post-hoc analyses using paired t-tests on session-to-session change scores were conducted within the ADHD group to explore significant group-by-time interaction effects. Clinical relevance was examined by correlating significant baseline voice features with parent-rated ADHD raw symptom scores. Additionally, linear regression models were used to test whether changes in ADHD symptoms (IA and HI) predicted changes in voice features, adjusting for age and gender. Given the small sample size and the exploratory nature of the study, no correction for multiple comparisons was applied.

Speech embeddings were analysed using LMMs largely analogous to those applied to voice features. Due to the high dimensionality of the embeddings (768 dimensions per model), task-specific *Principal Component Analysis* (PCA) was conducted on baseline data. For each task, the first ten principal components (PCs) were retained and used for subsequent analyses. Using a fixed number of components ensured comparable dimensions while maintaining consistency for the subsequent analysis. For each of the two encoder models, 10 PCs were entered as fixed effects in the LMM alongside baseline age and gender. This resulted in 60 separate LMMs for each encoder (10 PCs by six tasks).

To assess whether observed effects of speech embeddings were independent of established voice features, PCs showing significant baseline or interaction effects were reanalysed including specific voice features as covariates (mean F0, standard deviation of MFCC1, and standard deviation of voiced-segment length).

3 Results

3.1 Demographic and clinical characteristics

Table 2 shows demographic and clinical characteristics of the 54 included participants. The groups did not differ significantly in age ($t = 1.14, p = .260$) or gender ($\chi^2 = 2.08, p = .353$) distributions at baseline. ADHD symptoms (Conners 3-P) significantly decreased from baseline to follow-up in the ADHD group (Inattention (IA): $t(23) = 5.63, p < .001$; Hyperactivity/Impulsivity (HI): $t(23) = 3.98, p = .001$), whereas no significant change was observed in controls (IA: $t(23) = -1.85, p = 0.08$; HI: $t(23) = 0.13, p = 0.90$). 47 participants had voice data available for both visits (see **Appendix S5** for information on missing sessions). The mean interval between visits was 11.93 weeks (SD = 6.18) for children with ADHD and 8.42 weeks (SD = 0.82) for controls. The longer and more variable interval in the ADHD group reflects differences in the timing of pharmacotherapy initiation among drug-naïve participants. Details on recording equipment, the number of completed tasks per visit, and the average duration of preprocessed recordings per task are provided in **Table S1**.

Medication

Children with ADHD were unmedicated at the baseline visit. Twenty-two patients were initially drug-naïve and prescribed MPH according to routine clinical care. Titration, timing of initiation, and dosage were individualised in line with current guidelines. Three participants had paused prior MPH treatment before baseline (treatment duration before study entry: 5 years, 10 months, 4 months; pause duration before follow-up: 41, 44 and 51 days).

Following the baseline assessment, MPH treatment was initiated/resumed according to the individual treatment plan. Participants were expected to receive continuous treatment for approximately eight weeks until follow-up. At follow-up, children with ADHD were assessed while taking their individually prescribed MPH medication. Twenty-three children received an extended-release formulation, one an immediate-release formulation, and one a combination of both (**Table S2** for details). Voice recordings were conducted on medication, on average 3 h 51 min (SD = 1 h 13 min) after the last intake.

3.2 Voice features

3.2.1 Baseline differences

Six unique voice features across three tasks showed significant baseline group differences after FDR correction (**Table 3, Figure 1**). These features were predominantly pitch-related and included measures of F0, capturing both mean level and variability. The values of these features indicated that children with ADHD had lower voices that varied more during speech and therefore seemed less stable. Spectral features also contained discriminative information, particularly the first and second Mel-frequency cepstral coefficients (MFCCs), which characterize spectral shape such as voice timbre. These parameters showed that voices of children with ADHD were brighter/tenser (MFCC1), and less variable in timbre (MFCC2), meaning the tone colour fluctuated less. Additionally, the standard deviation of voiced-segment length, an indicator of rhythmic stability, differentiated between the groups, suggesting that the voices of children with ADHD were less rhythmically stable. Residuals of the LMM were approximately normally distributed. Robustness checks using ranked-transformed data to mitigate the influence of outliers yielded mostly consistent results with respect to effect estimates and signs. Comparable results were obtained when excluding the three children with ADHD who were assessed during a medication pause. Additionally, various features displayed significant group differences not surviving correction for multiple comparisons (**Table S3**).

The picture description task, followed by fiction and retell tasks, appeared to best elicit baseline differences. Features derived from the picture description task yielded the largest absolute β -values, indicating generally stronger effects, resulting in several significant voice features under both uncorrected ($n = 5$) and FDR-corrected ($n = 5$) conditions.

3.2.2 Interaction effects

Two voice features from separate tasks exhibited significant group-by-time effects after FDR correction, indicating that these parameters changed differently between visits in children with ADHD compared to controls (**Table 4, Figure 2**). Notably, the picture description task revealed the strongest interaction effect: The mean bandwidth of the first formant (F1) increased in the ADHD group, largely converging with the comparably stable control values. This indicates a more pronounced change in vowel articulation for children with ADHD and is consistent with clearer articulation over time. The derived results of the fiction task showed that the ADHD group exhibited decreased loudness variability, suggesting a more stable energy pattern over time, while the control group showed greater variability. Residuals were approximately normally distributed. The interaction effects

remained stable in ranked-data LMMs and while excluding the three children with paused medication. Additionally post-hoc sensitivity analyses with adjustment for the time between visits, that differed between the two groups, produced similar results, with the group \times time interaction effects remaining unchanged in terms of direction and significance. Features not surviving correction for multiple comparisons are reported in **Table S4**.

Post-hoc analysis on the change scores within the ADHD group confirmed significant changes over time with moderate within-group effect sizes (Bandwidth F1: $t(23) = 3.54, p = .002, dz = 0.72$; Loudness: $t(14) = -2.30, p = .038, dz = -0.59$).

3.2.3 Clinical Relevance

Correlation voice features and ADHD symptom ratings

Associations between baseline voice features and parent-rated ADHD symptoms were generally small and not statistically significant. However, mean f_0 (F0), showed a moderate positive correlation with hyperactivity/impulsivity (HI) values ($r = .407, p = .048$) consistent with the direction of the baseline group differences identified in the LMM analyses.

Linear Regression Model using change scores

Linear regression analyses were conducted using voice features that showed a robust interaction effect (**Table 4**). Within the ADHD group, changes in inattentive (IA) symptoms were significantly associated with changes in the mean bandwidth of the first formant (F1) ($\beta = -0.032, SE = 0.014, p = .034$). This indicates that a greater reduction in IA symptoms corresponds to larger increase in F1 bandwidth and thus more pronounced vowel articulation. Changes in HI symptoms were not associated with F1 bandwidth ($p = .272$). For loudness variability, neither symptom dimension showed significant associations with voice changes (IA: $p = .523$; HI: $p = .230$).

3.3 Speech Embeddings

Separate LMMs for each PC revealed significant baseline and group-by-time interaction effects for both WavLM and Whisper embeddings, after adjustment for baseline age, gender and selected voice features [f_0 mean, MFCC1 standard deviation (SD), voiced-segment length (SD)]. For the WavLM model, two interaction effects remained significant after FDR correction; the other effects did not survive multiple comparison analysis (**Table S5**). Whisper embeddings showed four baseline effects and one interaction effect that survived FDR correction (**Table S6**). Taken together, these findings suggest that speech embeddings capture variance not accounted for by theoretically defined voice features and may provide complementary information on ADHD-control differences

at baseline and medication-related changes over time. Across tasks, the first 10 principal components explained on average 76.82% (WavLm) / 65.57% (Whisper) of the variance. Explained variance for each PC and task are reported in **Table S7** (WavLM) and **Table S8** (Whisper).

4 Discussion

This study represents a proof-of-concept investigation of voice-based markers in a clinically characterised ADHD sample. The primary aim was to evaluate the suitability of voice features to distinguish between children with ADHD and controls and to capture treatment-related changes. Our main findings demonstrate that voice features robustly capture both group differences between children with ADHD and neurotypical controls and potential medication-related changes in children with ADHD, in a task-dependent manner. Exploratory analyses using speech embeddings suggest that the speech signal may contain additional information beyond established voice features, motivating more targeted investigations of, e.g., (para)linguistic features in future studies.

Several voice features differed significantly in children with ADHD compared to controls. Unmedicated children with ADHD had a lower, more variable voice pitch. At the same time, the colour of their voices seemed to be brighter and more tense, which could be perceived as pressed. This is in line with previous findings, describing a lower but more strained voice in children with ADHD [13, 14]. A strained voice may result from excessive vocal effort, which can be caused by altered motor regulation accompanied by heightened muscle tone in ADHD [38]. Furthermore, children with ADHD appear to exhibit more phonotraumatic behaviour, which involves placing greater mechanical stress on the vocal folds [39]. From a neuropsychological perspective, hypoarousal [40] may weaken top-down regulatory control necessary for vocal motor output and pitch regulation, contributing to increased pitch variability. Children with ADHD spoke less rhythmically during the counting task, which involved counting from one to ten. Counting is repetitive and therefore has a rhythm of its own. Children with ADHD may struggle to generate an internal beat [41] and have difficulties with the precise timing of motor tasks [42], which may explain this finding. We also found that the voice in both groups changed differently over time. After medication, children with ADHD showed a clearer articulation and a more stable loudness pattern. ADHD has been linked to changes in the brain's dopamine signalling [43], which could affect voice production [44]. The medication's effects include altering central dopamine activity [45], which could positively influence speech production. Controls displayed a reversed pattern for loudness variability, a speculative interpretation is increased comfort or familiarity at follow-up, though this cannot be determined from the current data. For articulation precision, the values almost merged from the first to the second visit. Given the naturalistic design, the effects of medication and time cannot be fully disentangled. However, the absence of comparable changes in the control group suggests that the changes observed are medication related.

We found multiple significant baseline and interaction effects for the speech embeddings. Given their limited interpretability, especially following necessary dimensionality reduction, we avoid a discussion of the meaning of significant embeddings. However, their training objectives could provide indirect clues about the types of speech information they encode (predominantly acoustic-prosodic vs. linguistic-phonetic). While our results demonstrate the potential of embedding-based analysis of speech data in ADHD, future research should examine which (para)linguistic aspects are captured by embeddings. This is crucial as underlying neuropsychological processes, such as emotional self-regulation and impulsivity [46] as well as inhibitory control [11], may contribute to altered (para)linguistic patterns in ADHD. Further exploration of the capability of the two models to capture baseline (Whisper) vs. longitudinal (WavLM) alterations would also be valuable. Combining acoustic features with speech embeddings in machine learning models may further improve performance [47].

The final stated goal of this study was to examine speech tasks in terms of their ability to discriminate between children with ADHD and controls, to provide prior knowledge for future task design. The picture description task yielded the most robust features, likely because children produce similar content, while differing subtly in prosodic delivery. This dual benefit makes description tasks suitable to elicit connected speech, which can provide clinical information as previous research has shown [48, 49]. The fiction and retell tasks also appear suitable. They require the children to either generate content freely and think outside of the box, or to retrieve, hold and manipulate prior knowledge to verbalise it. These tasks require cognitive resources enabled by e.g. executive functions (EF) like cognitive flexibility and working memory [50]. Speech and articulation also rely on coordinated neural networks [6] and higher order EF [51]. Reduced cognitive or attentional resources might hinder stable integration of articulatory patterns, and increased cognitive load could alter speech-motor control [6]. In the fiction and retell tasks, children with ADHD might allocate more cognitive resources to content generation than to articulation or prosody. The three tasks above fall into the category of *free speech*, highlighting the necessity of the inclusion of such tasks, as is already common in speech- and voice-based research [21].

Importantly, speech-based markers were evaluated in relation to an established clinical ADHD diagnosis and should therefore be interpreted as distinguishing clinically diagnosed from neurotypical children, rather than as stand-alone diagnostic markers. This study has several limitations. First, the modest sample size may have limited the detection of subtle effects, particularly in the linear regression analysis and correlations between voice features and symptom-scores. Second, recording equipment was updated during data collection; intensity-dependent features from the earlier setup were therefore excluded. Finally, as this was an observational study in children with ADHD, several aspects could not be standardised or monitored, like the time between visits, medication

timing, or daily behaviour/feelings (like sleep or stress). Strengths of this study include a clinically well characterized ADHD cohort, most of whom were drug-naïve at baseline, and a well-matched control group with repeated assessments. Additionally, high-quality voice recordings were obtained in accordance with current guidelines and a standardised study protocol.

In summary, to our knowledge, this is the first clinical study to systematically examine task-dependent voice features and speech embeddings in children with ADHD, including their sensitivity to pharmacological treatment effects of MPH. The findings demonstrate that speech-based measures capture both disorder-related vocal characteristics and medication-associated change patterns in children with ADHD, with the picture description task being particularly informative.

Acknowledgements

The authors would like to thank the families and children/adolescents for their participation. They also thank Lena Siepe, Katharina Ebi and Franziska Laven for their valuable support to this study especially during data acquisition.

Ethical Information

The study is in accordance with the Declaration of Helsinki, was approved by the *Ethics Committee of the Medical Faculty of RWTH Aachen University* (approval no. 032/20), and registered with the *German Clinical Trials Registry* (DRKS; registration no. DRKS00030766). Written informed consent was received from all parents and children, regardless of age.

Funding

Some components of the overarching study are part of the ABCD-J project (“Accessing Behavior for Clinical Data and Joint Usage (ABCD-J): A Platform for Behavioral Markers of Digital Neuromedicine in North Rhine-Westphalia”), which is funded by the *Ministry of Culture and Science of the State of North Rhine-Westphalia* (MKW NRW; cooperation platforms 2022, grant number KP22-106A). The study was funded by the START-Program of the *Faculty of Medicine RWTH Aachen University* which included funding for R.B.’s salary. L.D.L was supported by the Federal Ministry of Education and Research (BMBF) and the Max Planck Society (MPG), Germany.

Contributions

R.B. and G.P. conceptualised the study. R.B., G.K., G.P. designed the SOPs for the voice assessments. R.B conducted the data acquisition with the help of research assistants. R.B., G.P., L.D.L, J.D. performed the statistical analysis. R.B. wrote the manuscript. All authors were involved in critically revising the manuscript, and all authors approved the final version.

Declaration of interests

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

Data/Code availability

The raw data underlying this study cannot be shared publicly due to the sensitive and potentially identifiable nature of voice recordings and associated clinical data. Where ethically and legally permissible, derived and non-identifiable summary features may be shared upon reasonable request and subject to data use agreements. We provide the full analysis code in a GitHub repository (<https://github.com/RaBa-fzj/adhd-voice-features>)

Corresponding author

Correspondence to Georg von Polier (g.von.polier@fz-juelich.de) and Rachel Bamberger (r.bamberger@fz-juelich.de)

6 References

1. Wolraich ML, Hagan JF, Jr., Allan C, Chan E, Davison D, Earls M, et al. Clinical Practice Guideline for the Diagnosis, Evaluation, and Treatment of Attention-Deficit/Hyperactivity Disorder in Children and Adolescents. *Pediatrics*. 2019 Oct;144(4). PMID: 31570648. doi: 10.1542/peds.2019-2528.
2. Hurjui IA, Hurjui RM, Hurjui LL, Serban IL, Dobrin I, Apostu M, et al. Biomarkers and Neuropsychological Tools in Attention-Deficit/Hyperactivity Disorder: From Subjectivity to Precision Diagnosis. *Medicina*. 2025;61(7):1211. PMID: doi:10.3390/medicina61071211.
3. Eom TH, Kim YH. Clinical practice guidelines for attention-deficit/hyperactivity disorder: recent updates. *Clinical and Experimental Pediatrics*. 2024 Jan;67(1):26-34. PMID: 37321571. doi: 10.3345/cep.2021.01466.
4. Wong ICK, Banaschewski T, Buitelaar J, Cortese S, Döpfner M, Simonoff E, et al. Emerging challenges in pharmacotherapy research on attention-deficit hyperactivity disorder-outcome measures beyond symptom control and clinical trials. *Lancet Psychiatry*. 2019 Jun;6(6):528-37. PMID: 31122482. doi: 10.1016/s2215-0366(19)30096-3.
5. Emser TS, Johnston BA, Steele JD, Kooij S, Thorell L, Christiansen H. Assessing ADHD symptoms in children and adults: evaluating the role of objective measures. *Behavioral and Brain Functions*. 2018 2018/05/18;14(1):11. doi: 10.1186/s12993-018-0143-x.
6. Bouchard KE, Mesgarani N, Johnson K, Chang EF. Functional organization of human sensorimotor cortex for speech articulation. *Nature*. 2013 2013/03/01;495(7441):327-32. doi: 10.1038/nature11911.
7. Kappen M, Vanderhasselt MA, Slavich GM. Speech as a promising biosignal in precision psychiatry. *Neuroscience and Biobehavioral Reviews*. 2023 May;148:105121. PMID: 36914080. doi: 10.1016/j.neubiorev.2023.105121.
8. Fagherazzi G, Fischer A, Ismael M, Despotovic V. Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice. *Digital Biomarkers*. 2021 Jan-Apr;5(1):78-88. PMID: 34056518. doi: 10.1159/000515346.
9. Low DM, Bentley KH, Ghosh SS. Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope Investigative Otolaryngology*. 2020 Feb;5(1):96-116. PMID: 32128436. doi: 10.1002/liv.2.354.
10. Etter NM, Cadely FA, Peters MG, Dahm CR, Neely KA. Speech motor control and orofacial point pressure sensation in adults with ADHD. *Neuroscience Letters*. 2021 2021/01/23;744:135592. doi: <https://doi.org/10.1016/j.neulet.2020.135592>.
11. Engelhardt PE, Corley M, Nigg JT, Ferreira F. The role of inhibition in the production of disfluencies. *Memory & Cognition*. 2010 2010/07/01;38(5):617-28. doi: 10.3758/MC.38.5.617.
12. Barona-Lleo L, Fernandez S. Hyperfunctional Voice Disorder in Children With Attention Deficit Hyperactivity Disorder (ADHD). A Phenotypic Characteristic? *Journal of Voice*. 2016 Jan;30(1):114-9. PMID: 25862604. doi: 10.1016/j.jvoice.2015.03.002.
13. Garcia-Real T, Diaz-Roman TM, Garcia-Martinez V, Vieiro-Iglesias P. Clinical and acoustic vocal profile in children with attention deficit hyperactivity disorder. *Journal of Voice*. 2013 Nov;27(6):787 e11-8. PMID: 24246332. doi: 10.1016/j.jvoice.2013.06.013.

14. Hamdan AL, Deeb R, Sibai A, Rameh C, Rifai H, Fayyad J. Vocal characteristics in children with attention deficit hyperactivity disorder. *Journal of Voice*. 2009 Mar;23(2):190-4. PMID: 18082369. doi: 10.1016/j.jvoice.2007.09.004.
15. Bamberger R, Lotter LD, Nieto N, Poulain T, Körner A, Kiess W, et al. Voices of change: associations between vocal markers and symptoms of ADHD - Findings from the LIFE child study. *European Child & Adolescent Psychiatry*. 2026 2026/02/26. doi: 10.1007/s00787-025-02954-9.
16. von Polier GG, Ahlers E, Volkening J, Langner J, Patil KR, Eickhoff SB, et al. Exploring voice as a digital phenotype in adults with ADHD. *Scientific Reports*. 2025 2025/05/24;15(1):18076. doi: 10.1038/s41598-025-01989-x.
17. Bloch Y, Aviram S, Neeman R, Braw Y, Nitzan U, Maoz H, et al. Methylphenidate mediated change in prosody is specific to the performance of a cognitive task in female adult ADHD patients. *World Journal of Biological Psychiatry*. 2015;16(8):635-9. PMID: 25945954. doi: 10.3109/15622975.2015.1036115.
18. Congologlu MA, Turkbay T, Ciyiltepe M, Durukan I, Dogangun B, Yuce M. Immediate effects of methylphenidate on vocal acoustic parameters in children with attention deficit hyperactivity disorder. *Verus Journal*. 2009;19(4):365-72. doi: [https://www.psychiatry-
psychopharmacology.com/index.php/pub/article/view/569](https://www.psychiatry-psychopharmacology.com/index.php/pub/article/view/569).
19. Adnan T, Abdelkader A, Liu Z, Hossain E, Park S, Islam MS, et al. A novel fusion architecture for detecting Parkinson's Disease using semi-supervised speech embeddings. *NPJ Parkinsons Disease*. 2025 Jun 20;11(1):176. PMID: 40541966. doi: 10.1038/s41531-025-00956-7.
20. Maji B, Nasreen S, Guha R, Routray A, Majumdar D, Poonam K. Exploring Self-Supervised Models for Depressive Disorder Detection: A Study on Speech Corpora. *IEEE Engineering in Medicine and Biology Society*. 2024 Jul;2024:1-4. PMID: 40039538. doi: 10.1109/embe53108.2024.10781765.
21. Hecker P, Steckhan N, Eyben F, Schuller BW, Arnrich B. Voice analysis for neurological disorder recognition—a systematic review and perspective on emerging trends. *Frontiers in Digital Health*. 2022;4:842301. doi: 10.3389/fdgth.2022.842301.
22. Föcking W, Parrino M. *Praxis der Funktionalen Stimmtherapie: 3+ 1 Konzept*©: Springer; 2022.
23. Kaufman J, Birmaher B, Axelson D, Perepletchikova F, Brent D, Ryan N. *K-SADS-PL DSM-5* Yale University. 2016.
24. Lidzba K, Christiansen H, Drechsler R. *Conners Skalen Zu Aufmerksamkeit Und Verhalten—3 (Conners 3®)*: Hogrefe; 2013.
25. Beushausen U, Bohlender JE, Brockmann-Bauser M, Caffier P, Dicks P, Deuster D, et al. Diagnostik und Therapie von Störungen der Stimmfunktion (Dysphonien). *S2k-Leitlinie (049-008)*. 2022.
26. Patel RR, Awan SN, Barkmeier-Kraemer J, Courey M, Deliyski D, Eadie T, et al. Recommended Protocols for Instrumental Assessment of Voice: American Speech-Language-Hearing Association Expert Panel to Develop a Protocol for Instrumental Assessment of Vocal Function. *American Journal of Speech-Language Pathology*. 2018 Aug 6;27(3):887-905. PMID: 29955816. doi: 10.1044/2018_ajslp-17-0009.
27. Brockmann M, Drinnan MJ, Storek C, Carding PN. Reliable jitter and shimmer measurements in voice clinics: the relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical clinical task. *Journal of Voice*. 2011 Jan;25(1):44-53. PMID: 20381308. doi: 10.1016/j.jvoice.2009.07.002.

28. Eyben F, Scherer KR, Schuller BW, Sundberg J, André E, Busso C, et al. The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Transactions on Affective Computing*. 2015;7(2):190-202.
29. Eyben F, Wöllmer M, Schuller B. Opensmile: the munich versatile and fast open-source audio feature extractor. *Proceedings of the 18th ACM international conference on Multimedia*; Firenze, Italy: Association for Computing Machinery; 2010. p. 1459–62.
30. Chen S, Wang C, Chen Z, Wu Y, Liu S, Chen Z, et al. Wavlm: Large-scale self-supervised pre-training for full stack speech processing. *IEEE Journal of Selected Topics in Signal Processing*. 2022;16(6):1505-18. doi: <https://arxiv.org/abs/2110.13900>.
31. Radford A, Kim JW, Xu T, Brockman G, McLeavey C, Sutskever I, editors. Robust speech recognition via large-scale weak supervision. *International Conference on Machine Learning*; 2023: PMLR.
32. McFee B, Matt McVicar, Daniel Faronbi, Iran Roman, Gover M, Stefan Balke, et al. Librosa (0.11.0). Zenodo2025.
33. Jadoul Y, Thompson B, de Boer B. Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*. 2018 2018/11/01/;71:1-15. doi: <https://doi.org/10.1016/j.wocn.2018.07.001>.
34. Šrámková H, Granqvist S, Herbst CT, Švec JG. The softest sound levels of the human voice in normal subjects. *Journal of Acoustical Society America*. 2015 Jan;137(1):407-18. PMID: 25618070. doi: 10.1121/1.4904538.
35. Berger T, Peschel T, Vogel M, Pietzner D, Poulain T, Jurkutat A, et al. Speaking Voice in Children and Adolescents: Normative Data and Associations with BMI, Tanner Stage, and Singing Activity. *Journal of Voice*. 2019 Jul;33(4). PMID: 29807693. doi: 10.1016/j.jvoice.2018.01.006.
36. Lavan N, Burton AM, Scott SK, McGettigan C. Flexible voices: Identity perception from variable vocal signals. *Psychonomic Bulletin & Review*. 2019 2019/02/01;26(1):90-102. doi: 10.3758/s13423-018-1497-7.
37. Rashid MM, McKean JW, Kloke JD. Review of rank-based procedures for multicenter clinical trials. *Journal of Biopharmaceutical Statistics*. 2013;23(6):1207-27. PMID: 24138428. doi: 10.1080/10543406.2013.834919.
38. Stray LL, Stray T, Kristensen O. Neuromuscular regulation problems in attention deficit hyperactivity disorder assessed by the motor function neurological assessment. *Journal Novel Physiotherapies*. 2015;5(255):2. doi: 10.4172/2165-7025.1000255.
39. Beşik Topçu Ö, Esen Aydınlı F, Kirazlı MÇ, Öğülmüş Uysal M, Aslan C, Çak Esen HT, et al. A Descriptive Study on the Vocal Behaviors in Children With Attention Deficit Hyperactivity Disorder: Vocal Hygiene Knowledge and Phonotraumatic Behaviors. *Journal of Voice*. 2025 2025/05/15/. doi: <https://doi.org/10.1016/j.jvoice.2025.04.028>.
40. Bellato A, Arora I, Hollis C, Groom MJ. Is autonomic nervous system function atypical in attention deficit hyperactivity disorder (ADHD)? A systematic review of the evidence. *Neuroscience & Biobehavioral Reviews*. 2020 2020/01/01/;108:182-206. doi: <https://doi.org/10.1016/j.neubiorev.2019.11.001>.
41. Puyjarinet F, Bégel V, Lopez R, Dellacherie D, Dalla Bella S. Children and adults with Attention-Deficit/Hyperactivity Disorder cannot move to the beat. *Scientific Reports*. 2017 2017/09/14;7(1):11550. doi: 10.1038/s41598-017-11295-w.
42. Zelaznik HN, Vaughn AJ, Green JT, Smith AL, Hoza B, Linnea K. Motor timing deficits in children with Attention-Deficit/Hyperactivity disorder. *Human Movement Science*. 2012 2012/02/01/;31(1):255-65. doi: <https://doi.org/10.1016/j.humov.2011.05.003>.

43. Dum R, Ghahramani A, Baweja R, Bellon A. Dopamine Receptor Expression and the Pathogenesis of Attention-Deficit Hyperactivity Disorder: a Scoping Review of the Literature. *Current Developmental Disorders Reports*. 2022 2022/12/01;9(4):127-36. doi: 10.1007/s40474-022-00253-5.
44. Turk AZ, Lotfi Marchoubeh M, Fritsch I, Maguire GA, SheikhBahaei S. Dopamine, vocalization, and astrocytes. *Brain and Language*. 2021 2021/08/01;219. doi: <https://doi.org/10.1016/j.bandl.2021.104970>.
45. Faraone SV. The pharmacology of amphetamine and methylphenidate: Relevance to the neurobiology of attention-deficit/hyperactivity disorder and other psychiatric comorbidities. *Neuroscience and Biobehavioral Reviews*. 2018 Apr;87:255-70. PMID: 29428394. doi: 10.1016/j.neubiorev.2018.02.001.
46. Faraone SV, Rostain AL, Blader J, Busch B, Childress AC, Connor DF, et al. Practitioner Review: Emotional dysregulation in attention-deficit/hyperactivity disorder—implications for clinical recognition and intervention. *Journal of child psychology and psychiatry*. 2019;60(2):133-50.
47. Li S, Nair R, Naqvi SM. Acoustic and Text Features Analysis for Adult ADHD Screening: A Data-Driven Approach Utilizing DIVA Interview. *IEEE Journal of Translational Engineering in Health and Medicine*. 2024;12:359-70. PMID: 38606391. doi: 10.1109/jtehm.2024.3369764.
48. Mueller KD, Hermann B, Mecollari J, Turkstra LS. Connected speech and language in mild cognitive impairment and Alzheimer's disease: A review of picture description tasks. *Journal of Clinical and Experimental Neuropsychology*. 2018 2018/10/21;40(9):917-39. doi: 10.1080/13803395.2018.1446513.
49. Patel R, Connaghan K. Park Play: A picture description task for assessing childhood motor speech disorders. *International Journal of Speech-Language Pathology*. 2014 2014/08/01;16(4):337-43. doi: 10.3109/17549507.2014.894124.
50. Diamond A. Executive functions. *Annual Review of Psychology*. 2013;64(1):135-68.
51. Netelenbos N, Gibb RL, Li F, Gonzalez CLR. Articulation Speaks to Executive Function: An Investigation in 4- to 6-Year-Olds. *Frontiers in Psychology*. 2018 2018-February-26;Volume 9 - 2018. doi: 10.3389/fpsyg.2018.00172.
52. Berube S, Nonnemacher J, Demsky C, Glenn S, Saxena S, Wright A, et al. Stealing Cookies in the Twenty-First Century: Measures of Spoken Narrative in Healthy Versus Speakers With Aphasia. *American Journal of Speech-Language Pathology*. 2019;28(1S):321-9. doi: doi:10.1044/2018_AJSLP-17-0131.

7 Tables and Figures

Table 1 Name, content and target recording length of the six speech tasks for both visits

Task Name	Content	Target recording length (s)
T1: Frequency	State the own name and count from 1-10	Variable
T2: Picture description	Description of the <i>Cookie Theft</i> picture [52]	90
T3: Retell	Description of something nice that has already happened (e.g. memorable day, school trip)	60
T4: Fiction	Description of the best day possible, if everything was possible without limitations (e.g. by money)	60
T5: Phonation normal	Hold and phonate /a/ three times for 5 seconds each in comfortable pitch and loudness	9
T6: Phonation loud	Hold and phonate /a/ three times for 5 seconds each in a louder voice than before	9

Note. For all tasks beside *Phonation loud* children were instructed to speak in the intensity and pitch they normally do and feel comfortable with. T = Task, s = seconds.

Table 2 Demographic and clinical information for both visits

Characteristic	Group	n (%) (Baseline)	Mean/SD (Baseline)	n (Follow-Up)	Mean/SD (Follow-Up)
Age	ADHD	27	12.79 (2.65)	25	13.04 (2.77)
	Control	27	12.01 (2.38)	27	12.17 (2.38)
Gender					
Male	ADHD	10	-	-	-
	Control	11	-	-	-
Female	ADHD	15	-	-	-
	Control	16	-	-	-
Divers	ADHD	2	-	-	-
	Control	0	-	-	-
ADHD Subtype					
Combined	ADHD	18	-	-	-
Inattentive		8	-	-	-
Hyperactive-Impulsive		1	-	-	-
Conners3-P					
Inattention	ADHD	25	17.72 (5.52)	24	11.79 (6.65)
	Control	26	4.12 (3.28)	25	5.24 (4.13)
Hyperactivity/ Impulsivity	ADHD	25	17.20 (10.05)	24	11.04 (8.59)
	Control	26	4.08 (4.13)	25	3.88 (4.48)
Comorbid diagnosis					
Depressive Disorders	ADHD	22 (81.48)	-	-	-
Anxiety Disorders		13 (48.15)	-	-	-
Trauma-related Disorders		12 (44.44)	-	-	-
Gender Dysphoria		4 (14.81)	-	-	-
Eating Disorders		2 (7.41)	-	-	-
Specific Learning Disorder		2 (7.41)	-	-	-
Somatic Symptom Disorder		2 (7.41)	-	-	-
Conduct Disorders		1 (3.70)	-	-	-
Communication Disorders		1 (3.70)	-	-	-
Elimination Disorders		1 (3.70)	-	-	-

Note. SD = Standard deviation; Conners3-P = Conners 3rd Edition-Parent (long form). Comorbid diagnoses were limited to psychiatric and behavioural disorders classified in the DSM-5. Comorbid diagnoses were not mutually exclusive, meaning that children could meet the criteria for multiple additional diagnoses.

Table 3 FDR-corrected significant voice features at baseline (linear mixed model)

Task	Feature Name	ADHD (n)	Control (n)	ADHD Mean	Control Mean	df_resid	coef (β)	se	ci_lower	ci_upper	p	q
Picture	F0_mean	26	26	-0.514	0.304	93	-0.629	0.215	-1.051	-0.207	0.003	0.038
	Description											
Picture	F0_SDnorm	26	26	1.046	0.436	93	0.581	0.172	0.244	0.917	0.001	0.031
	Description											
Picture	F0_p20	26	26	-0.871	0.148	93	-0.907	0.301	-1.496	-0.318	0.003	0.032
	Description											
Picture	MFCC2_SDnorm	26	26	-0.077	0.069	93	-0.146	0.047	-0.239	-0.053	0.002	0.031
	Description											
Picture	MFCC1V_SDnorm	26	26	0.950	0.480	93	0.470	0.133	0.209	0.732	0.000	0.031
	Description											
Fiction	MFCC1V_SDnorm	25	26	0.871	0.429	93	0.452	0.130	0.197	0.708	0.001	0.046
Frequency	VoicedSeg_Len_SD	26	26	0.181	-0.212	93	0.401	0.110	0.186	0.616	0.000	0.023

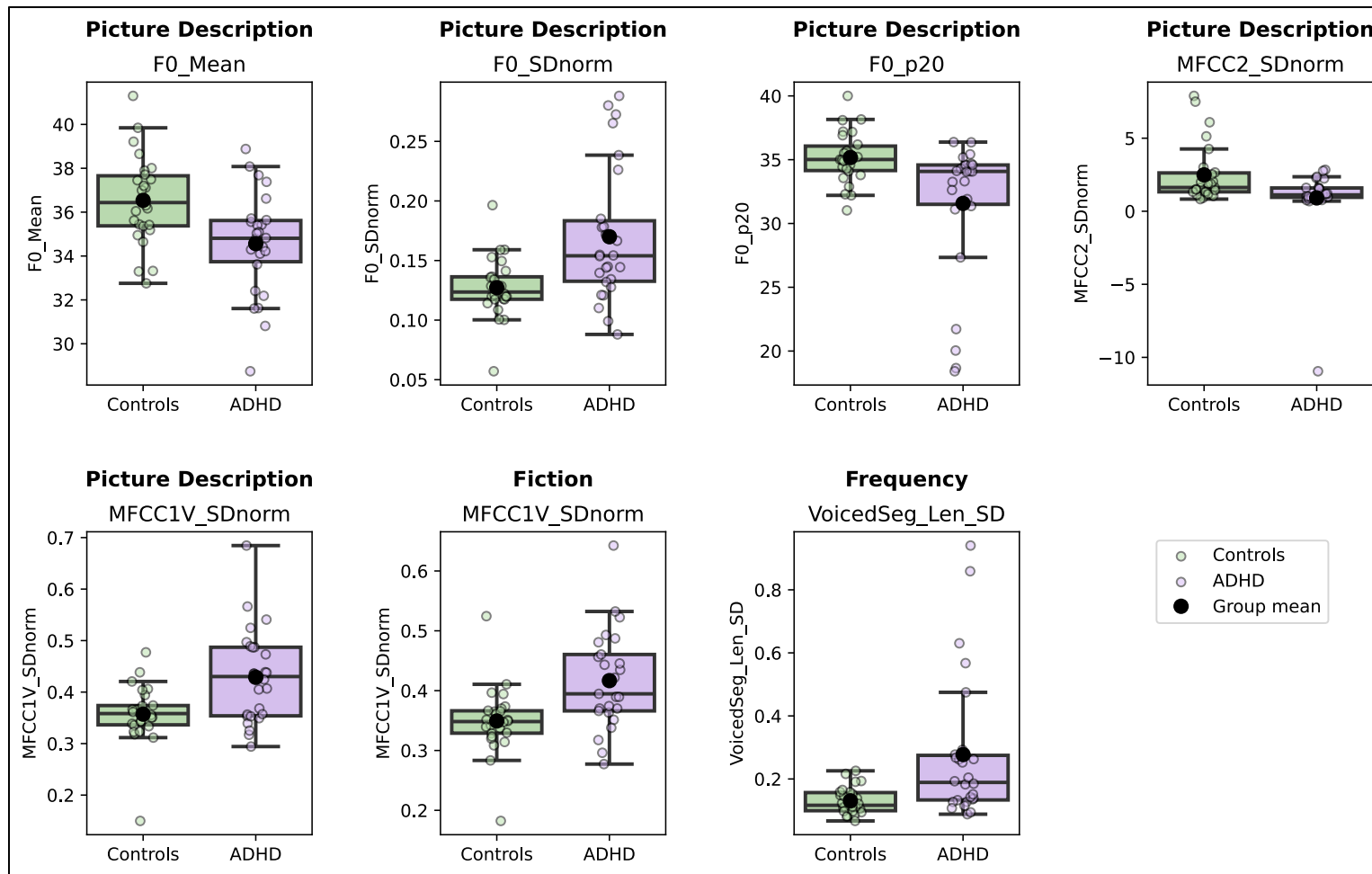
Note. All models were adjusted for the covariates baseline age and gender; all models converged successfully; reported degrees of freedom refer to the model residuals and reflect the total number of repeated observations included in each mixed-effects model; df_resid = degrees of freedom (residual); coef = coefficient; se = standard error; ci_lower/upper = confidence interval (lower/upper bound); p/q = p-/q-value (FDR adjusted p-value); F0_mean = mean pitch (fundamental frequency); F0_SDnorm = normalized pitch variability; F0_p20 = 20th percentile pitch; MFCC2_SDnorm = normalized variability of 2nd Mel-frequency cepstral coefficient; MFCC1V_SDnorm = normalized variability of 1st MFCC in voiced regions; VoicedSeg_Len_SD = standard deviation of voiced segment lengths.

Table 4 FDR-corrected significant Group \times Time interaction effects in voice features (linear mixed model)

Task	Feature Name	ADHD Baseline/ Follow-Up	Control Baseline/ Follow-Up	ADHD Mean Baseline/ Follow-Up	Control Mean Baseline/ Follow-Up	df_resid	coef (β)	se	ci_lower	ci_upper	p	q
Picture Description	F1_bandwidth_mean	26/24	26/24	-0.393/-0.219	-0.134/-0.201	93	0.276	0.078	0.123	0.429	0.000	0.037
Fiction	Loudness_SDnorm	17/15	24/23	0.684/0.582	0.527/0.716	72	-0.353	0.097	-0.543	-0.163	0.000	0.024

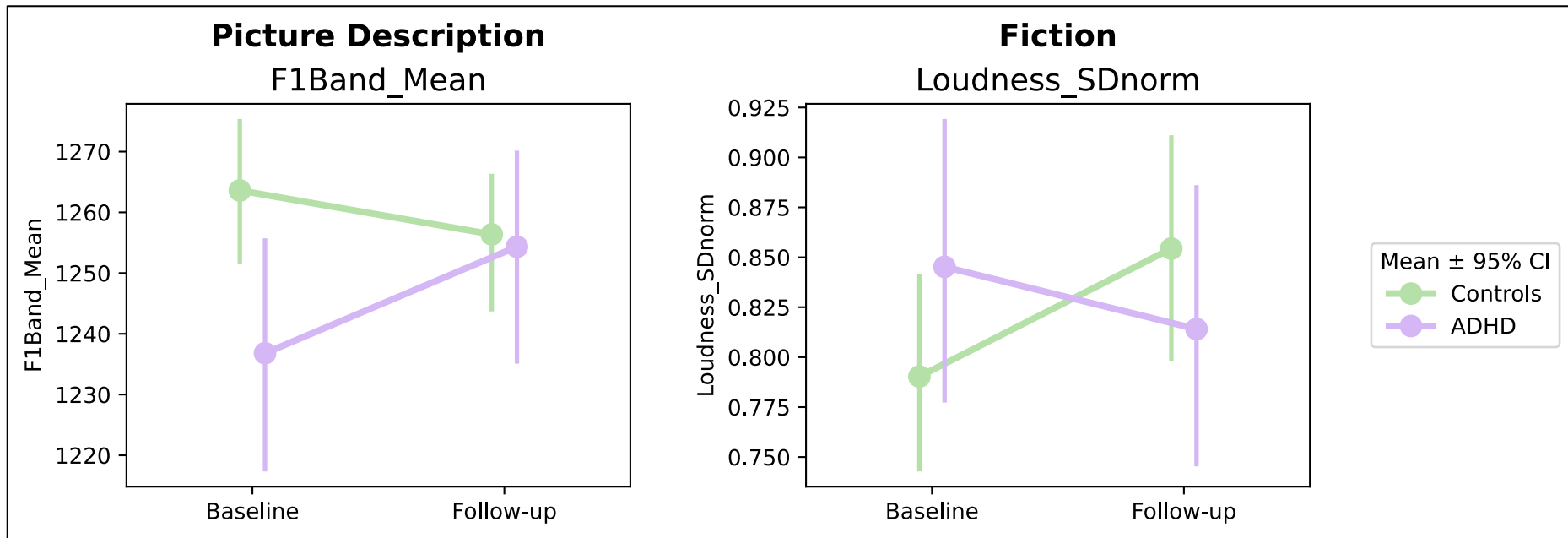
Note. All models were adjusted for the covariates baseline age and gender; all models converged successfully; df_resid = degrees of freedom (residual); coef = coefficient; se = standard error; ci_lower/upper = confidence interval (lower/upper bound); p/q = p-/q-value (FDR adjusted p-value); F1_bandwidth_mean = mean bandwidth of the first formant; Loudness_SDnorm = normalized variability of loudness.

Figure 1 FDR-corrected significant voice features at baseline (linear mixed model)



Note. All values represent the raw values of the significant effects found in the linear mixed model analysis. F0_Mean = mean pitch (fundamental frequency); F0_SDnorm = normalized pitch variability; F0_p20 = 20th percentile pitch; MFCC2_SDnorm = normalized variability of 2nd Mel-frequency cepstral coefficient; MFCC1V_SDnorm = normalized variability of 1st MFCC in voiced regions; VoicedSeg_Len_SD = standard deviation of voiced segment lengths.

Figure 2 Raw values of FDR-corrected significant Group \times Time interaction effects in voice features (linear mixed model)



Note. All values represent the raw values of the significant effects found in the linear mixed model analysis. Children with ADHD were under medication at follow-up.

F1Band_mean = mean bandwidth of the first formant; Loudness_SDnorm = normalized variability of loudness; CI = confidence interval.