RESEARCH

Open Access

The voice of depression: speech features as biomarkers for major depressive disorder

Felix Menne[1*], Felix Dörr[1], Julia Schräder[2,3], Johannes Tröger[1], Ute Habel[2,3], Alexandra König[1,4] and Lisa Wagels[2,3]

Abstract

Background Psychiatry faces a challenge due to the lack of objective biomarkers, as current assessments are based on subjective evaluations. Automated speech analysis shows promise in detecting symptom severity in depressed patients. This project aimed to identify discriminating speech features between patients with major depressive disorder (MDD) and healthy controls (HCs) by examining associations with symptom severity measures.

Methods Forty-four MDD patients from the Psychiatry Department, University Hospital Aachen, Germany and fifty-two HCs were recruited. Participants described positive and negative life events, which were recorded for analysis. The Beck Depression Inventory (BDI-II) and the Hamilton Rating Scale for Depression gauged depression severity. Transcribed audio recordings underwent feature extraction, including acoustics, speech rate, and content. Machine learning models including speech features and neuropsychological assessments, were used to differentiate between the MDD patients and HCs.

Results Acoustic variables such as pitch and loudness differed significantly between the MDD patients and HCs (effect sizes $\boldsymbol{\eta}2$ between 0.183 and 0.3, $p < 0.001$). Furthermore, variables pertaining to temporality, lexical richness, and speech sentiment displayed moderate to high effect sizes ($\boldsymbol{\eta}2$ between 0.062 and 0.143, $p < 0.02$). A support vector machine (SVM) model based on 10 acoustic features showed a high performance (AUC = 0.93) in differentiating between HCs and patients with MDD, comparable to an SVM based on the BDI-II (AUC = 0.99, $p = 0.01$).

Conclusions This study identified robust speech features associated with MDD. A machine learning model based on speech features yielded similar results to an established pen-and-paper depression assessment. In the future, these findings may shape voice-based biomarkers, enhancing clinical diagnosis and MDD monitoring.

Keywords Precision psychiatry, Depression, Speech biomarkers, Machine learning

*Correspondence:
Felix Menne
felix.menne@ki-elements.de
[1] ki:elements GmbH, Bleichstr. 27, 66111 Saarbrücken, Germany
[2] Department of Psychiatry, Psychotherapy and Psychosomatics, RWTH Aachen University, Aachen, Germany
[3] Institute of Neuroscience and Medicine: JARA-Institute Brain Structure Function Relationship (INM 10), Research Center Jülich, Jülich, Germany
[4] Université Côte d'Azur, Centre Hospitalier et Universitaire, Clinique Gériatrique du Cerveau et du Mouvement, Centre Mémoire de Ressources et de Recherche, Nice, France

Background

Major depressive disorder (MDD) is one of the most prevalent psychiatric conditions, with varying prevalence rates across regions, affecting up to approximately 10% of the population [1]. The economic burden of MDD was estimated at $326.2 billion for 2018 in the US alone [2]. Additionally, the condition is linked to lasting quality of life deficits even after remission [3] and persisting disability [4]. MDD stems from various factors, including genetic, biological, environmental, and psychological influences [5]. Among other symptoms, the disorder is characterised by persistent feelings of sadness, hopelessness, and a lack of interest or pleasure in daily activities [6]. Treatment options for MDD often involve a combination of psychotherapy, medication, and lifestyle changes [7].

Compared to other areas of medicine such as neurology, where objective biomarkers are well-established [8], psychiatry significantly lags behind, often relying on subjective assessments by patients and clinicians. However, recent advancements have demonstrated, for instance, that integrating neuroimaging, genetic, and clinical predictors through machine learning has enabled the prediction of therapeutic outcomes in depression patients with an accuracy of 0.82 [9]. A growing number of multimodal digital biomarkers have emerged to objectively assess behavioural or biological information for psychiatric conditions [10, 11]. Among these, speech analysis presents significant opportunities for studying disease-related characteristics [12], as psychiatric symptoms often manifest in speech and language. Speech has been recognised as a potential target in the context of predicting self-harm, suicidal behaviour, substance abuse, depression, and disease recurrence [13]. The relevant speech patterns might include speech rate, coherence, and content for various psychiatric conditions, such as depression, schizophrenia, or posttraumatic stress disorder [14–17]. Advances in computational linguistics, natural language processing, and speech recognition have facilitated the use of automatic speech analysis as an objective clinical measurement of psychiatric symptoms [18, 19].

Natural speech tasks can effectively elicit emotional responses by asking participants to describe events that recently triggered strong emotions. Unlike simple vocal exercises or reading tasks, these tasks can capture the acoustic effects of emotional changes [20]. Using emotion-induced speech tasks provides a wider range of emotional responses, such as recounting events that elicited significant emotions [21]. Specifically, temporal, prosodic and spectral features have been consistently reported to be associated with depression [22, 23]. These include speech characterised by monotony and flatness, i.e., a perception of being "lifeless". This has been attributed to reductions in the fundamental frequency f0 and the f0 range [24, 25]. Additionally, MDD patients often exhibit reduced speech rates and utterance durations, possibly linked to psychomotor retardation [26, 27]. This association was shown for both depression severity and treatment response, mainly for temporal features such as longer pause times or slower speaking rates [26]. Moreover, deviations in voice quality markers, such as jitter and shimmer, as well as spectral features, have been observed in depressed individuals [28–31]. Linguistic changes have also been documented, including heightened self-referential speech and increased use of past tense verbs [15, 32]. Furthermore, depressed individuals tend to express more emotionally negative content and employ less complex vocabulary [33, 34]. Additionally, there is substantial evidence supporting the classification of depressive syndrome severity using speech biomarkers, whether through dimensional or categorical approaches [35, 36]. Significant speech features identified in such studies include temporal aspects [27], voice arousal [37], and language features [38]. This evidence demonstrates the relevance of speech changes in depression patients.

To envision the use of speech and language markers in regular clinical practice, validation against gold standard measures is essential. This study aimed to explore the differences in speech features between clinical and healthy subjects, assess the impact of depression on patients' speech, and examine how speech characteristics relate to symptom severity measures. Based on previous findings, we expect that certain speech features related to frequency, temporal aspects, voice quality, and content will effectively distinguish between healthy controls (HCs) and depressed patients. Additionally, we postulate that temporal speech features will differentiate between mildly and moderately depressed participants. Additionally, we hypothesise that a classification model incorporating selected speech features will outperform a baseline model based solely on demographic and clinical characteristics in distinguishing between the two groups.

Methods

Participants

The participants recruited for this study took part in an investigation of unconscious emotional conflict in patients with MDD [39]. Patients diagnosed with MDD were recruited at the Department of Psychiatry, Psychotherapy and Psychosomatics, RWTH Aachen University, Germany. Age- and sex-matched HCs were recruited through advertisements and flyers in Aachen. Demographic information was assessed with a questionnaire.

The internal ethics committee of the university hospital Aachen, Germany, approved the study (Ethik-Kommission an der Medizinischen Fakultät der RWTH Aachen; vote number EK 045−19). All study procedures were conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants prior to participation in the study. All participants received 45 Euros of financial compensation.

Clinical assessment

Using a clinical interview and the German version of the Structured Clinical Interview based on the DSM-V (SCID-V [40]) the diagnostic assessments were conducted by trained psychologists.

On the day of the study visit, the severity of symptoms was assessed based on self-reports using the Beck Depression Inventory (BDI-II [41]) as well as medication status, clinical assessment and interviews using the Hamilton Depression Rating Scale (HAMD, 21-item version [42]). The classification of patients as mildly or moderately depressed was based on the number of clinical symptoms according to the DSM-5 criteria. A diagnosis of mild depression requires the presence of five symptoms, while moderate depression is indicated by six to seven symptoms [40]. The total score on the HAMD was used as a supplementary tool to assist in determining the severity of depression.

Procedure

The data were collected over a period of 18 months at the Department of Psychiatry, Psychotherapy, Psychosomatics, Medical Faculty RWTH Aachen University. The included participants were invited to a 3.5-hour measurement appointment.

Participants underwent a series of neuropsychological tests and psychopathological rating scales before performing two tasks in a 3-tesla MRI scanner while simultaneously recording electroencephalography. The findings on this paradigm are presented in Schräder et al., 2024 [39], whereas this paper focuses on speech as an exploratory secondary analysis. The neuropsychological tests included the Trail-Making-Test A/B (TMT A/B [43]) to assess processing speed and executive functioning. Furthermore, the Digit Span (DS) forwards and backwards test, a subtest of the Wechsler Memory Scale (WMS-R [44]), was conducted to measure short-term and working memory. Additionally, clinical and mood variables were assessed through the State-Trait Anxiety Inventory (STAI [45]). Afterwards, the participants completed the speech task.

Healthy controls were identified by combining data from the clinical interview and questionnaire scores. If there were indications of a clinically significant psychiatric or somatic condition, participants were advised to seek appropriate clinical care. This included, but was not limited to, conditions affecting voice and speech, such as reading disabilities, speech delays or vocal cord disorders.

Speech assessment and processing

Participants were asked to talk about both a positive and a negative event in their lives, a paradigm used in previous studies to elicit emotion in speech [46, 47]. The instructions ("Can you tell me in one minute about a positive/negative event in your life?"; see also Supplementary file for the original German version) were pre-recorded by a psychologist and played from a tablet, ensuring consistent instructions across both experiments. The responses were recorded using the tablet's internal microphone. To compute the acoustic and linguistic features, the extraction scripts were implemented in Python 3.9, based on our own speech processing library "Sigma". The extraction code is available upon reasonable request.

We investigated several categories of acoustic and linguistic speech features that were previously associated with symptoms of depression [15, 26, 27, 29, 34, 48]. Speech attributes were categorised into groups, sorted distinctly by negative and positive story (refer to Table 1 for groupings and corresponding attributes).

These included acoustic components such as frequency, energy and spectral features. The frequency features consist of variables associated with the so-called formants F0 to F3. F0 is commonly perceived as voice pitch [49]. The specific formants F1, F2 and F3 make up the expression of vowel sounds [50]. Energy features pertain to attributes such as jitter, which reflects voice instability, shimmer, which indicates changes in voice loudness or intensity, and the harmonics-to-noise ratio (HNR [51]). Spectral features describe characteristics of speech frequencies, such as the mel frequency cepstral coefficient (MFCC), a measure used for speaker identification [52].

Additionally, we employed features delineated by our research team concerning the temporal dimensions of speech [47], which encompass timing-related traits of verbal expression, such as duration, rhythm, and temporal patterns [53].

Moreover, we used linguistic attributes, including pronouns, adjectives, adverbs, and conjunctions. These categories, termed lexical richness, syntactic complexity, and word types, were defined by our research team [54].

Finally, to capture emotional response, we explored sentiment, i.e. positive or negative emotional tones. For this purpose, we used an external Python library called Stanza [55]. This library taps into extensive language models, particularly neural networks employing contextualised word embeddings. Preexisting language models identify word categories and determine whether sentences convey positivity, neutrality, or negativity. Stanza, an open-source Python natural language processing (NLP) toolkit, was developed by the Stanford NLP Group [56], supporting analysis across 66 diverse human languages.

Table 1 Categories of analysed categories with associated speech features

Feature group	Feature	Explanation
Energy		
	apq3_shimmer	Three-point period amplitude perturbation quotient (short-term amplitude variability in vocal fold vibrations).
	apq5_shimmer	Five-point period amplitude perturbation quotient.
	apq11_shimmer	Eleven-point period amplitude perturbation quotient.
	dda_shimmer	Dynamic Decline Amplitude Shimmer (average absolute differences between the amplitudes of consecutive periods)
	hnr_mean	Mean Harmonic-to-Noise Ratio (HNR) in decibels, calculated using the cepstral analysis of the acoustic signal.
	hnr_sd	Standard deviation of hnr_mean.
	local_shimmer	Cycle-to-cycle variability in amplitude (as percentage of the average amplitude).
	loudness_mean	Mean speech loudness.
	loudness_sd	Standard deviation of loudness_mean.
	rate_loudness_peaks	Frequency of loudness peaks in a given time period.
	shimmer_local_dB_mean	Cycle-to-cycle variability in amplitude in decibels.
	shimmer_local_dB_sd	Standard deviation of difference of shimmer_local_dB_mean.
Frequency		
	ddp_jitter	Dynamic Decline Perturbation in jitter (average absolute differences between the fundamental frequencies of consecutive periods).
	f0_range	Range of the fundamental frequency (F0) of vocal fold vibrations.
	f1_bandwidth_mean	Mean bandwidth of F1 formant.
	f1_bandwidth_sd	Standard deviation of f1_bandwidth_mean.
	f1_frequency_mean	Mean frequency of F1 formant.
	f1_frequency_sd	Standard deviation of f1_frequency_mean.
	f2_bandwidth_mean	Mean bandwidth of F2 formant.
	f2_bandwidth_sd	Standard deviation of f2_bandwidth_mean.
	f2_frequency_mean	Mean frequency of F2 formant.
	f2_frequency_sd	Standard deviation of f2_frequency_mean.
	f3_bandwidth_mean	Mean bandwidth of F3 formant.
	f3_bandwidth_sd	Standard deviation of f3_bandwidth_mean.
	f3_frequency_mean	Mean frequency of F3 formant.
	f3_frequency_sd	Standard deviation of f3_frequency_mean.
	jitter_local_mean	Deviations in individual consecutive F0 period lengths (perceived as uneven or irregular voice).
	jitter_local_sd	Standard deviation of jitter_local_mean.
	local_absolute_jitter	Cycle-to-cycle variability in the fundamental frequency of vocal fold vibrations.
	pitch_coefficient_of_variation	Coefficient of variation of pitch.
	pitch_first_quartile	First quartile of pitch.
	pitch_kurtosis	Degree of peakedness or flatness of the fundamental frequency (F0) distribution.
	pitch_linear_regression_mse	Mean squared error of a linear regression model applied to fundamental frequency (F0) values.
	pitch_linear_regression_offset	Intercept of a linear regression model applied to fundamental frequency (F0) values.
	pitch_linear_regression_slope	Slope coefficient of a linear regression model applied to the fundamental frequency (F0) values.
	pitch_max	Maximum pitch.
	pitch_mean	Mean pitch.
	pitch_min	Minimum pitch.
	pitch_percentile_1	First percentile of pitch.
	pitch_percentile_20	20th percentile of pitch.
	pitch_percentile_20_80_range	Range of 20th to 80th percentile of pitch.
	pitch_percentile_80	80th percentile of pitch.

Table 1 (continued)

Feature group	Feature	Explanation
	pitch_percentile_99	99th percentile of pitch.
	pitch_percentile_1_99_range	Range of 1st to 99th percentile of pitch.
	pitch_q2_q1_range	Range of 1st to 2nd quartile of pitch.
	pitch_q3_q1_range	Range of 1st to 3rd quartile of pitch.
	pitch_q3_q2_range	Range of 2nd to 3rd quartile of pitch.
	pitch_range	Range of pitch.
	pitch_second_quartile	Second quartile of pitch.
	pitch_skewness	Asymmetry of the fundamental frequency (F0) distribution.
	pitch_std	Standard deviation of pitch.
	pitch_third_quartile	Third quartile of pitch.
	ppq5_jitter	Five-point period perturbation quotient, quantifying the cycle-to-cycle variability in the fundamental frequency (F0) of vocal fold vibrations.
	rap_jitter	Relative Average Perturbation jitter, to quantify the cycle-to-cycle variation in the fundamental frequency (F0) of vocal fold vibrations.
	vocal_tremor	Vocal tremor, measuring the intensity of low-frequency F0 modulation, defined as the peak magnitude of F0 modulation within the 1.5–15 Hz band.
Lexical Richness		
	brunets_index	Brunet's Index, measure of lexical diversity.
	honore_stat	Honoré's statistic, measure to assess the lexical richness or diversity.
	number_consecutive_repetitions	Number of consecutive repetitions.
	type_token_ratio	Proportion of unique words (types) to total words (tokens) in a text or speech sample.
	word_count	Number of words used.
	word_frequency_mean	Mean occurrence rate of each word token, calculated as the mean frequency of all tokens across the utterance.
	word_frequency_range	Range of word_frequency_mean.
	word_frequency_sd	Standard deviation of word_frequency_mean.
Sentiment		
	mean_sentiment	Average emotional valence of the sentences (indicating if the whole answer was in general more positive, neutral, or negative).
	negative_sentence_ratio	Number of sentences that are labeled as negative in relation to all sentences.
	neutral_sentence_ratio	Number of sentences that are labeled as neutral in relation to all sentences.
	positive_sentence_ratio	Number of sentences that are labeled as positive in relation to all sentences.
Spectral		
	alpha_ratio_mean	Mean of the ratio of the summed energy from 50–1000 Hz and 1–5 kHz.
	alpha_ratio_sd	Standard deviation of alpha_ratio_mean.
	average_mfccs_1	Average of the Mel-Frequency-Cepstral-Coefficient 1 (Measurement to capture fundamental frequency characteristics of human speech).
	average_mfccs_2	Average of the Mel-Frequency-Cepstral-Coefficient 2.
	average_mfccs_3	Average of the Mel-Frequency-Cepstral-Coefficient 3.
	average_mfccs_4	Average of the Mel-Frequency-Cepstral-Coefficient 4.
	f1_relative_energy_mean	Average relative energy of the first formant (F1), reflecting the prominence of F1 resonance.
	f1_relative_energy_sd	Standard deviation of f1_relative_energy_mean.
	f2_relative_energy_mean	Average relative energy of the second formant (F2), reflecting the prominence of F2 resonance.
	f2_relative_energy_sd	Standard deviation of f2_relative_energy_mean.
	f3_relative_energy_mean	Average relative energy of the third formant (F3), reflecting the prominence of F3 resonance.
	f3_relative_energy_sd	Standard deviation of f3_relative_energy_mean.
	h1_a3_harmonic_difference_mean	Mean of the ratio of energy of the first F0 harmonic (H1) to the energy of the highest harmonic in the third formant range (A3).
	h1_a3_harmonic_difference_sd	Standard deviation of h1_a3_harmonic_difference_mean.

Table 1 (continued)

Feature group	Feature	Explanation
	h1_h2_harmonic_difference_mean	Mean of the ratio of energy of the first F0 harmonic (h1) to the energy of the highest harmonic in the second formant range (H2).
	h1_h2_harmonic_difference_sd	Standard deviation of h1_h2_harmonic_difference_mean.
	hammarberg_index_mean	Ratio of energy between higher (2–5 kHz) and lower (0.5–2 kHz) frequency bands. Parameter to assess the spectral balance of a voice.
	hammarberg_index_sd	Standard deviation of Hammarberg Index.
	spectral_slope_0_500_mean	Mean of the linear regression slope of the logarithmic power spectrum within the two given bands.
	spectral_slope_0_500_sd	Standard deviation of spectral_slope_0_500_mean.
	spectral_slope_500_1500_mean	Mean of the linear regression slope of the logarithmic power spectrum within the two given bands.
	spectral_slope_500_1500_sd	Standard deviation of Slope V500-1500.
Syntactic Complexity		
	mean_number_subordinate_clauses	Number of subordinate clauses used.
	proportion_verb_phrase_with_objects	Proportion of phrases with the verb being directed towards an object (e.g. She ate a delicious meal).
	proportion_verb_phrase_with_subjects	Proportion of phrases with the verb being performed by the subject (e.g. She slept peacefully).
	verb_phrase_with_aux_and_vp_rate	Ratio of verb phrases that include both auxiliary verbs and main verbs (e.g. She has finished her homework).
	verb_phrase_with_aux_rate	Ratio of verb phrases that include an auxiliary verb.
Temporal		
	duration	Length of audio recording.
	length_continuously_unvoiced_regions_mean	Mean of continuously unvoiced regions.
	length_continuously_unvoiced_regions_sd	Standard deviation of length_continuously_unvoiced_regions_mean.
	length_continuously_voiced_regions_mean	Mean of continuously voiced regions.
	length_continuously_voiced_regions_sd	Standard deviation of length_continuously_voiced_regions_mean.
	number_of_pauses	Number of pauses in between speech segments based on speech intervals.
	pause_durations_mean	Mean length of pauses.
	pause_durations_sd	Standard deviation of pause duration.
	pause_durations_sum	Sum of pause lengths over whole utterance.
	pause_rate	Frequency of pauses within a spoken utterance.
	speech_ratio	Ratio of utterance to duration.
	utterance_durations_mean	Mean length of utterances.
	utterance_durations_sd	Standard deviation of utterance durations.
	utterance_durations_sum	Sum of utterance durations.
Word Types		
	adjective_rate	Relative frequency (ratio to total words spoken) of adjectives used.
	adposition_rate	Ratio of adpositions used.
	adverb_rate	Ratio of adverbs used.
	conjunction_rate	Ratio of conjunctions used.
	determiner_rate	Ratio of determiners (articles, demonstratives, possessives) used.
	inflected_verb_rate	Ratio of inflected verbs used.
	noun_rate	Ratio of nouns used.
	pronoun_rate	Ratio of pronouns used.
	proper_noun_rate	Ratio of proper nouns used.
	verb_rate	Ratio of verbs used.

Statistical analysis

The statistical analyses were conducted with the Python package scipy.stats (v1.11.4, Linux v5.10.0). For the analysis of demographic and clinical data, we reported non-parametric group comparisons between HC, mild and moderate depression via Kruskal-Wallis tests. Where the overall effect was significant, we examined the pairwise differences between groups by Mann-Whitney U tests.

To test for group differences in speech features between HCs and patients with MDD (both mildly and moderately depressed patients combined), Kruskal-Wallis H tests were conducted. $\boldsymbol{\eta}2$ serves as a measure of effect size. $\boldsymbol{\eta}2$ values of 0.01 are considered small, 0.06 medium and values > 0.14 are considered high effect sizes. Correlating speech features with age and TMT-A and conducting group comparisons with sex as outcome, revealed significant influences of sex and TMT-A (for detailed results on sex and TMT-A, see Supp. Tables 1 and 2). Thus, further analyses were controlled for sex to account for acoustic differences and TMT-A as a measure of potential psychomotor retardation [57].

Furthermore, we tested for group differences in speech features between the two groups with mild and moderate depression, controlling for age, sex, and TMT-A score.

To test for associations between speech features and the BDI-II score, Spearman rank sum correlations with Spearman's ρ as a measure of correlation were computed. ρ can take values between −1 and +1, whereas ±1 is considered a perfect positive/negative correlation. This analysis was conducted without stratifying for groups (i.e., across all participants, whether healthy or depressed). We controlled for age, sex, and TMT-A.

To examine the incremental utility of the selected speech features over a baseline model, classification models were created to discriminate between HCs and patients MDD and, additionally, between patients with mild and moderate depression. First, a "baseline model" consisting of age, sex and years of education as well as TMT A/B, DS Forwards/Backwards, and the MWT-B was used. Second, a BDI model including only the BDI-II questionnaire scores was used. The third model we defined was a speech model consisting of linguistic and acoustic features extracted from the story-telling task. To select features, we decided to sort the features based on their mutual information with the outcome and select the 10 best [58]. To avoid overfitting, we computed the performance metrics Balanced Accuracy and the Area under the Receiver-Operator Characteristic Curve (AUC) based on a 10-fold cross validation approach where the sample is split into 10 groups of equal size and one of the groups is left out for validation of the model. This process is repeated 10 times such that each group was once used for validation. To test for significant differences between the models, we computed pairwise permutation tests.

All *p*-values reported were adjusted for multiple hypothesis testing using the Benjamini-Hochberg procedure [59]. To do so, the features were clustered in categories as presented in Table 1 and the *p*-values were adjusted separately for each of the categories.

Results

A total of forty-four patients were included in the study. Of these, twenty-nine were defined as mildly depressed according to the clinical interview and the HAMD, and fifteen suffered from moderate depression. Fifty-two participants were defined as HCs. The demographic and clinical information of the participants can be found in Table 2.

Group differences in speech features between HCs and MDD patients

Several features emerged as significantly different between HCs and depressed patients (see Table 3). The greatest effect sizes were observed for the acoustic features. For instance, the pitch linear regression slopes for positive ($\boldsymbol{\eta}2 = 0.31$, adjusted $p < 0.001$) and negative ($\boldsymbol{\eta}2 = 0.30$, $p < 0.001$) storytelling. This variable reflects the intonation pattern of speech over time, i.e., a positive slope indicates that the pitch tends to rise over time, while a negative slope indicates a decreasing pitch.

Table 2 Demographic and clinical scores

	Total	Healthy Controls	Mild Depression	Moderate Depression	Kruskal-Wallis H	*p*-value	post-hoc-tests
n	**96**	**52**	**29**	**15**	-	-	-
Age	26.16 (6.65)	26.17 (7.08)	26.1 (5.85)	26.2 (7.0)	0.248	0.883	-
Years of education	13.29 (2.28)	14.06 (2.24)	12.24 (2.23)	12.67 (1.5)	12.371	**0.002**	HC > mild = moderate
% female	39.6	63.5	51.7	73.3	3.035	0.219	-
BDI-II	14.66 (13.53)	3.94 (3.13)	23.96 (9.18)	30.87 (11.31)	66.241	**< 0.001**	HC < mild = moderate
HAMD	19.84 (5.52)	-	16.55 (3.22)	25.33 (2.82)	30.84	**< 0.001**	mild < moderate
TMT-A	19.73 (5.02)	20.59 (5.52)	18.06 (4.23)	19.9 (3.86)	5.896	0.052	-
STAI-X1	39.67 (10.89)	32.53 (6.11)	46.04 (9.87)	50.13 (8.57)	42.09	**< 0.001**	HC < mild = moderate
STAI-X2	44.19 (14.57)	32.06 (6.63)	54.79 (5.78)	62.4 (8.18)	67.177	**< 0.001**	HC < mild < moderate
MWT-B	28.79 (3.1)	29.26 (2.7)	27.5 (3.81)	29.73 (2.02)	5.556	0.062	-
Digit Span Backwards	7.67 (2.61)	8.06 (2.63)	6.93 (2.58)	7.73 (2.46)	3.496	0.174	-
Digit Span Forwards	11.45 (2.55)	11.82 (2.44)	11.0 (2.64)	11.0 (2.7)	1.781	0.41	-

The values are given as the means ± standard deviations in brackets. Gender shares are given as percentages. The nonparametric Kruskal-Wallis H test was used to determine group differences. Post-hoc comparisons were conducted by Mann-Whitney U tests; BDI-II = Beck Depression Scale; HAMD = Hamilton Depression Rating Scale; TMT-A = Trail making test A; STAI X1/X2 = state-trait anxiety inventory sub-scales 1 and 2; MWT-B = Mehrfachwahl-Wortschatz-Intelligenztest B (multiple choice Vocabulary Intelligence Test version B)

Table 3 Group differences (Kruskal-Wallis H test) between HCs (0) and depressed patients [1]

	Total	0	1	H	*p*	*η*2	Adj. *p*
N	96	52	44				
pitch_linear_regression_slope_pos	-0.01	-0.02	-0.01	30.091	0.0	0.309	< 0.001
pitch_linear_regression_slope_neg	-0.01	-0.01	0.0	29.449	0.0	0.303	< 0.001
alpha_ratio_mean_pos	8.76	8.68	8.97	27.797	0.0	0.285	< 0.001
f3_relative_energy_mean_pos	-22.63	-21.88	-24.54	26.418	0.0	0.27	< 0.001
f3_relative_energy_mean_neg	-22.98	-22.57	-24.77	24.127	0.0	0.246	< 0.001
hammarberg_index_sd_neg	8.42	8.56	8.23	21.94	0.0	0.223	< 0.001
loudness_sd_pos	12.07	16.03	9.37	21.256	0.0	0.215	< 0.001
h1_h2_harmonic_difference_mean_neg	5.86	4.68	6.56	20.851	0.0	0.211	< 0.001
rate_loudness_peaks_neg	0.27	0.62	0.06	20.584	0.0	0.208	< 0.001
loudness_mean_pos	-51.09	-53.95	-48.36	20.318	0.0	0.206	< 0.001
alpha_ratio_mean_neg	9.11	8.49	9.57	20.185	0.0	0.204	< 0.001
loudness_sd_neg	12.02	17.07	9.31	19.725	0.0	0.199	< 0.001
length_continuously_unvoiced_regions_mean_neg	0.11	0.12	0.09	19.014	0.0	0.192	< 0.001
loudness_mean_neg	-51.83	-55.65	-49.11	18.567	0.0	0.187	< 0.001
apq5_shimmer_pos	8.1	7.45	9.75	17.691	0.0	0.178	< 0.001
rate_loudness_peaks_pos	0.26	0.46	0.06	17.322	0.0	0.174	< 0.001
apq11_shimmer_neg	13.53	12.01	17.35	16.356	0.0	0.163	< 0.001
apq5_shimmer_neg	8.41	7.43	9.91	16.237	0.0	0.162	< 0.001
length_continuously_unvoiced_regions_sd_neg	0.1	0.11	0.08	15.942	0.0	0.159	< 0.001
apq11_shimmer_pos	13.58	11.95	15.93	15.074	0.0	0.15	< 0.001
apq3_shimmer_neg	6.14	5.5	7.31	14.733	0.0	0.146	< 0.001
dda_shimmer_neg	18.41	16.51	21.94	14.733	0.0	0.146	< 0.001
apq3_shimmer_pos	6.17	5.62	7.03	14.341	0.0	0.142	< 0.001
dda_shimmer_pos	18.52	16.86	21.1	14.341	0.0	0.142	< 0.001
pitch_skewness_neg	0.18	0.4	0.07	17.567	0.0	0.176	0.001
f1_frequency_sd_neg	337.51	322.01	357.19	16.475	0.0	0.165	0.001
h1_h2_harmonic_difference_mean_pos	5.67	4.8	6.27	15.825	0.0	0.158	0.001
length_continuously_unvoiced_regions_mean_pos	0.11	0.12	0.1	15.592	0.0	0.155	0.001
pitch_kurtosis_neg	0.4	0.88	0.13	15.074	0.0	0.15	0.001
pitch_percentile_20_pos	98.43	96.1	101.09	14.79	0.0	0.147	0.001
utterance_durations_mean_pos	0.19	0.17	0.24	14.733	0.0	0.146	0.001
hammarberg_index_mean_pos	25.61	24.59	26.74	14.508	0.0	0.144	0.001
pitch_max_pos	299.85	321.75	273.75	14.564	0.0	0.144	0.001
pitch_percentile_80_pos	182.38	186.98	172.7	14.564	0.0	0.144	0.001
duration_neg	29.48	25.05	47.46	14.452	0.0	0.143	0.001
utterance_durations_sd_pos	0.13	0.11	0.17	14.452	0.0	0.143	0.001
brunets_index_neg	9.81	9.01	10.82	14.229	0.0	0.141	0.001
positive_sentence_ratio_neg	0.43	0.33	0.5	14.091	0.0	0.139	0.001
word_frequency_mean_neg	4.82	4.82	4.82	14.008	0.0	0.138	0.001
utterance_durations_mean_neg	0.19	0.18	0.22	13.463	0.0	0.133	0.001
utterance_durations_sum_neg	11.7	8.3	18.68	13.195	0.0	0.13	0.001
duration_pos	19.66	16.14	27.98	13.035	0.0	0.128	0.001
mean_sentiment_neg	0.33	0.22	0.5	12.824	0.0	0.126	0.001
hammarberg_index_sd_pos	8.44	8.79	7.89	12.876	0.0	0.126	0.001
utterance_durations_sd_neg	0.13	0.12	0.16	12.666	0.0	0.124	0.001
utterance_durations_sum_pos	8.06	6.37	12.1	12.251	0.0	0.12	0.001
shimmer_local_dB_sd_neg	1.37	1.37	1.37	11.995	0.001	0.117	0.001
pitch_percentile_20_neg	95.05	95.05	95.7	13.735	0.0	0.135	0.002
pitch_skewness_pos	0.28	0.54	0.03	13.302	0.0	0.131	0.002
hammarberg_index_mean_neg	25.75	24.51	27.39	12.824	0.0	0.126	0.002
f2_relative_energy_mean_neg	-17.42	-17.29	-17.67	11.691	0.001	0.114	0.002
pause_durations_sd_neg	0.31	0.31	0.33	10.707	0.001	0.103	0.002

Table 3 (continued)

	Total	0	1	H	*p*	*η*2	Adj. *p*
length_continuously_unvoiced_regions_sd_pos	0.1	0.11	0.09	10.707	0.001	0.103	0.002
hnr_mean_neg	-105.49	-105.92	-104.54	10.42	0.001	0.1	0.002
shimmer_local_dB_sd_pos	1.35	1.35	1.35	10.42	0.001	0.1	0.002
neutral_sentence_ratio_neg	0.4	0.5	0.28	10.184	0.001	0.098	0.002
pitch_max_neg	291.94	310.67	257.4	11.893	0.001	0.116	0.003
pitch_percentile_80_neg	179.08	186.3	166.57	11.893	0.001	0.116	0.003
honore_stat_neg	1785.46	1910.71	1689.43	10.611	0.001	0.102	0.003
pitch_kurtosis_pos	0.75	0.93	0.43	12.046	0.001	0.118	0.004
spectral_slope_500_1500_sd_neg	5.26	5.26	5.21	10.184	0.001	0.098	0.004
speech_ratio_neg	0.41	0.35	0.46	9.045	0.003	0.086	0.005
negative_sentence_ratio_pos	0.0	0.0	0.0	9.698	0.002	0.093	0.007
f2_relative_energy_mean_pos	-16.82	-16.56	-17.21	9.178	0.002	0.087	0.009
pause_rate_pos	0.46	0.48	0.41	7.973	0.005	0.074	0.009
pause_durations_sd_pos	0.24	0.23	0.24	7.89	0.005	0.073	0.009
f1_relative_energy_sd_pos	8.14	8.19	8.1	8.738	0.003	0.082	0.01
number_consecutive_repetitions_pos	0.0	0.0	0.0	9.629	0.002	0.092	0.012
brunets_index_pos	9.16	8.63	9.73	8.651	0.003	0.081	0.012
word_frequency_mean_pos	4.78	4.81	4.74	8.014	0.005	0.075	0.012
pause_rate_neg	0.48	0.5	0.47	6.969	0.008	0.063	0.013
average_mfccs_3_neg	18.25	15.07	20.6	7.726	0.005	0.072	0.014
spectral_slope_0_500_mean_neg	0.26	0.26	0.31	7.685	0.006	0.071	0.014
f3_frequency_mean_neg	2911.24	2945.12	2894.73	8.825	0.003	0.083	0.016
f1_frequency_sd_pos	325.89	315.88	333.83	8.913	0.003	0.084	0.017
honore_stat_pos	1698.73	1683.99	1721.88	6.891	0.009	0.063	0.017
hnr_mean_pos	-102.54	-104.31	-96.65	6.068	0.014	0.054	0.018
h1_a3_harmonic_difference_mean_neg	-20.28	-18.23	-22.21	6.853	0.009	0.062	0.019
speech_ratio_pos	0.43	0.4	0.49	6.287	0.012	0.056	0.019
f3_relative_energy_sd_pos	9.95	10.18	9.64	7.125	0.008	0.065	0.021
spectral_slope_500_1500_sd_pos	5.07	5.07	5.12	6.814	0.009	0.062	0.022
word_count_pos	47.0	39.0	64.0	6.068	0.014	0.054	0.022
pitch_percentile_99_pos	239.49	264.87	218.92	8.098	0.004	0.076	0.023
pitch_percentile_1_99_range_pos	167.98	193.61	151.81	7.89	0.005	0.073	0.023
jitter_local_mean_pos	0.04	0.04	0.04	7.767	0.005	0.072	0.023
length_continuously_voiced_regions_mean_pos	0.19	0.18	0.2	5.676	0.017	0.05	0.024
pitch_std_pos	41.54	47.35	37.34	7.442	0.006	0.069	0.025
average_mfccs_2_pos	11.04	9.78	11.69	6.141	0.013	0.055	0.027
h1_h2_harmonic_difference_sd_pos	7.67	7.87	7.1	6.104	0.013	0.054	0.027
length_continuously_voiced_regions_sd_pos	0.27	0.25	0.29	5.23	0.022	0.045	0.028
average_mfccs_2_neg	13.12	11.87	14.76	5.676	0.017	0.05	0.034
h1_a3_harmonic_difference_sd_neg	8.89	9.26	8.76	5.468	0.019	0.048	0.036
h1_a3_harmonic_difference_sd_pos	8.99	9.23	8.45	5.399	0.02	0.047	0.037

Variables are listed in descending order for effect size and adjusted *p* value. Pos=positive story; neg=negative story. H=measure for Kruskal-Wallis H test; *p*=unadjusted *p* value; ***η*2**=effect size; Adj. *p*=*p* value adjusted for multiple hypothesis testing according to the Benjamini-Hochberg procedure. For brevity, only features with adj. *p*<0.05 are reported here. The full list can be found in suppl. Table 3

Depressed participants showed a lower slope in the positive story compared to the HCs, but a greater slope in the negative story. Furthermore, we found significant features pertaining to the alpha ratio with effect sizes $\boldsymbol{\eta}2>0.20$ ($p<0.001$) for both negative and positive storytelling. The alpha ratio is defined as the ratio of spectrum intensity above and below 1000 Hz and is influenced by vocal loudness [60]. Additionally, the mean loudness of positive ($\boldsymbol{\eta}2=0.21$, $p<0.001$) and negative ($\boldsymbol{\eta}2=0.19$, $p<0.001$) storytelling significantly differed between groups, with depressed participants talking louder in both conditions.

There were differences in temporal features such as the duration of utterances (positive story: $\boldsymbol{\eta}2=0.15$, $p=0.001$; negative story: $\boldsymbol{\eta}2=0.13$, $p=0.001$), with depressed participants speaking longer. Additionally, differences were observed in the pause rate, defined as the total length of pauses divided by the total length of speech, including pauses. In our data, we noted a lower pause rate in depressed individuals for positive ($\eta2=0.07$, $p<0.01$) and negative stories ($\eta2=0.06$, $p<0.015$).

Furthermore, depressed participants used more words (mean=64) than HCs did (mean=39, $\eta2=0.54$, $p=0.02$) in the positive story. This trend was also evident in the negative story, although it did not achieve statistical significance (depressed mean=109, HC mean=52, $\eta2=0.01$, $p=0.6$). However, the difference in the number of words produced between positive and negative stories was more noticeable in depressed participants.

We observed significant differences in lexical richness between the groups. The Brunet's index (BI) indicates richer language with lower numbers [61]. According to our data, the BI was greater in depressed participants in the negative ($\eta2=0.14$ $p=0.001$) and the positive ($\eta2=0.08$ $p=0.012$) story.

Furthermore, we assessed speech sentiment, which evaluates whether the emotional tone of the words used is predominantly positive, negative, or neutral. In our analysis, we discovered several features that significantly differed between groups, such as the positive sentence ratio in the negative story. Depressed participants used more positive sentences (ratio=0.5) than HCs did (ratio=0.33, $\eta2=0.14$, adjusted $p=0.001$). This effect did not achieve statistical significance for the positive story (HC/depressed ratio=0.5, $p=0.7$). Additionally, differences were observed in the neutral sentence ratio ($\eta2=0.1$, adjusted $p=0.02$), with healthy participants employing more neutral sentences in negative storytelling. Again, this effect was not significant for the positive story ($p=0.7$).

Group differences in speech features between mildly and moderately depressed patients

Various features exhibited high effect sizes when comparing mildly and moderately depressed individuals. However, none of the effects reached statistical significance (α<0.05). Among the variables with high effect sizes, most were related to pauses, such as pause rate, number of pauses, and pause duration ($\eta2>0.093$). Moderately depressed individuals showed more and longer pauses than those with mild depression. Additionally, those with moderate depression spoke at a lower volume than individuals with mild depression ($\eta2=0.093$). Variables related to voice quality also displayed differences, such as shimmer (indicating irregularities in the loudness of the voice), with moderately depressed individuals showing higher values than those with mild depression ($\eta2=0.066$). All these listed effects were derived from the negative, not the positive story. For a detailed breakdown, the full results are available in Supplementary Table 4.

Correlations between speech features and the BDI-II

Different speech features exhibited low to moderate correlations with the BDI-II (Fig. 1). For instance, variables pertaining to the MFCCS, both in negative and positive storytelling, demonstrated correlations (*r*) between 0.32 and 0.4, $p<0.001$, respectively. Furthermore, features such as jitter (irregularity in pitch, which can manifest

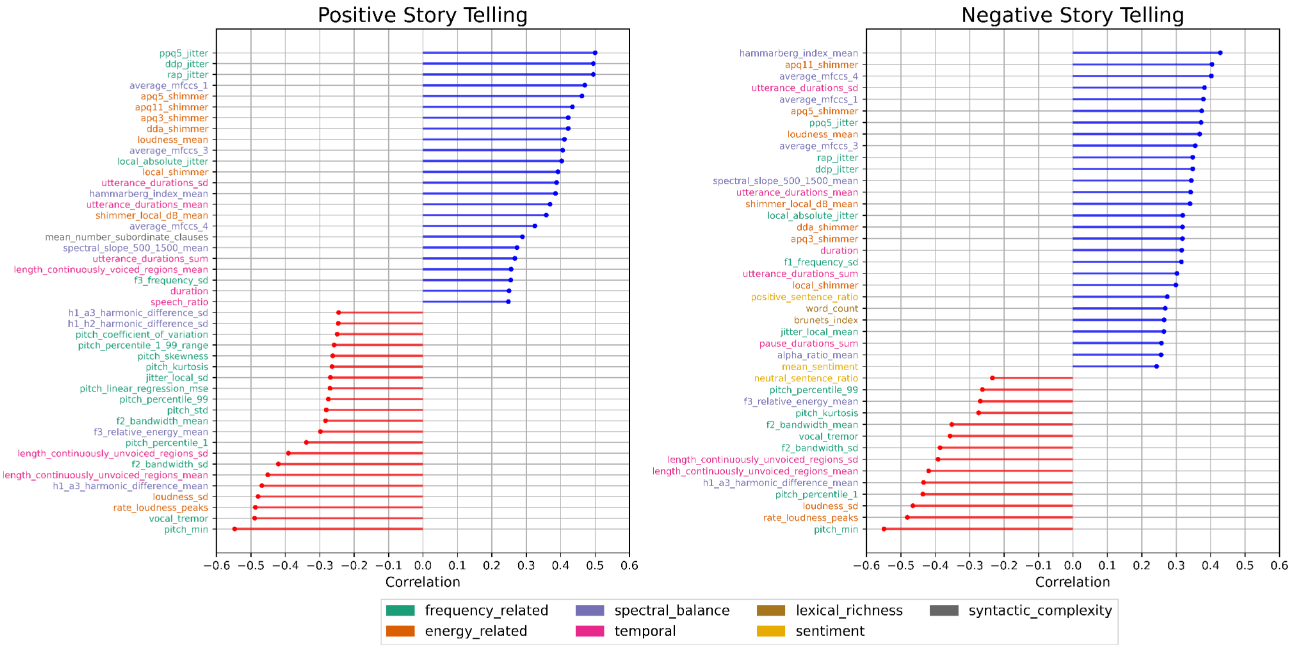

Fig. 1 Lollipop plots depicting results of Spearman rank sum correlations (*r*) between speech features and the BDI-II. Variables are colour-coded according to the captions, based on the categories defined in Table 1. For brevity, only features with adj. $p<0.05$ are depicted here. All results including effect sizes and *p* values can be found in Supp. Table 5

as a wavering or unstable quality in the voice; $r > 0.51$, $p < 0.001$, respectively) in the positive story and shimmer (irregularity in amplitude, leading to fluctuations in loudness or intensity; $r > 0.40$, $p < 0.001$, respectively) in both the positive and negative stories showed moderate correlations. Another feature showing a significant correlation was loudness, both in positive ($r = 0.40$, $p < 0.001$) and negative storytelling ($r = 0.35$, $p < 0.001$). The highest correlations were found for the minimal pitch in positive (r=-0.53, $p < 0.001$) and negative (r=-0.54, $p < 0.001$) storytelling. The complete list of correlation results can be found in Supp. Table 5.

Classification models

For each of the six classifications, several models were computed to select the one with the best performance with regard to the Balanced Accuracy (BA) and receiver operating characteristics - area under the curve (ROC-AUC). The models tested were random forest, extra trees, support vector machines (SVMs), linear models (LMs) and decision trees (DTs).

For detailed results of the comparison of HCs versus patients with MDD, refer to Table 4; Fig. 2. The best performing speech model consisted of 10 features, 7 of which were selected over all LOOCV iterations (rate_loudness_peaks_neg, vocal_tremor_neg, pitch_min_neg, average_mfccs_1_pos, apq5_shimmer_pos, h1_a3_harmonic_difference_mean_pos, pitch_min_pos). Notably, all seven features included over all iterations were identified as significant in the group comparisons and correlations ($p < 0.05$, respectively).

For detailed results of the comparison of mild vs. moderately depressed patients, refer to Table 4; Fig. 3. The speech DT model consisted of 50 features (for feature names and number of selections by means of LOOCV, see Supp. Table 6) and had a BA of 0.68 and an AUC of 0.69. In this instance, 15 of the 50 features included in the model were identified as significant in the group comparisons and correlations. Of these, two features—rate_loudness_peaks_neg and pitch_min_neg—were also part of the SVM model to distinguish between depressed and healthy participants.

The results of the permutation tests used to assess differences between ROC curves are presented in Table 5. For the HC vs. MDD models, there was no significant difference between the BDI-II and the speech model. However, both models significantly differed from the baseline models.

Discussion

In this research, we aimed to assess speech characteristics in a sample of depressed individuals and healthy controls. We examined group differences and correlations between specific speech features and BDI-II scores. Additionally, we developed classification models to differentiate between depressed and healthy participants based on speech features that show similar group discrimination effectiveness compared to classical measures such as the BDI-II. Our findings suggest that various speech features are significantly associated with depression when compared to healthy individuals.

The most prominent group differences between HCs and patients with MDD were demonstrated in variables pertaining to pitch slope, both for the positive and negative stories. For the negative story, we found a slope of zero in depressed participants (compared to -0.01 in HCs, $p < 0.001$), which is in line with existing data indicating a flatter pitch slope in depressed patients than in HCs [23]. Additionally, we found further variables related to pitch, such as kurtosis and skewness, to be lower in depressed participants than in healthy controls, both in the negative and the positive story conditions. Low kurtosis suggests a flatter distribution, indicating a more consistent pitch over the whole of the recording. Negative skewness indicates a longer left tail of the audio signal, meaning more instances of lower pitch values. These data indicate a generally flatter and monotonous speech in the depressed sample.

Furthermore, our analyses demonstrated the strongest correlations between depressive symptoms (BDI-II) and the minimum pitch in both the negative and positive stories. These results are generally in line with the literature demonstrating lower pitch variability and pitch slope in depressed individuals than in HCs [23, 29, 62]. A more recent study, however, demonstrated contrasting results,

Table 4 Performance metrics for different classification models

Comparison	Model	k features selected	BA	ROC-AUC	Sensitivity	Specificity
HC vs. MDD	Baseline model (LM)	9	0.62	0.72	0.56	0.68
	Speech Model (SVM)	10	0.88	0.93	0.81	0.94
	BDI-II Model (SVM)	1	0.94	0.99	0.91	0.98
Mild vs. moderate MDD	Baseline model (DT)	9	0.66	0.64	0.47	0.86
	Speech Model (LM)	50	0.56	0.62	0.33	0.79
	BDI-II Model (XT)	1	0.61	0.55	0.40	0.82

BA = balanced accuracy; DT = decision trees; LM = Linear Model; ROC-AUC = Receiver operating characteristic - area under the curve; SVM = support Vector Machine; XT = Extra Trees

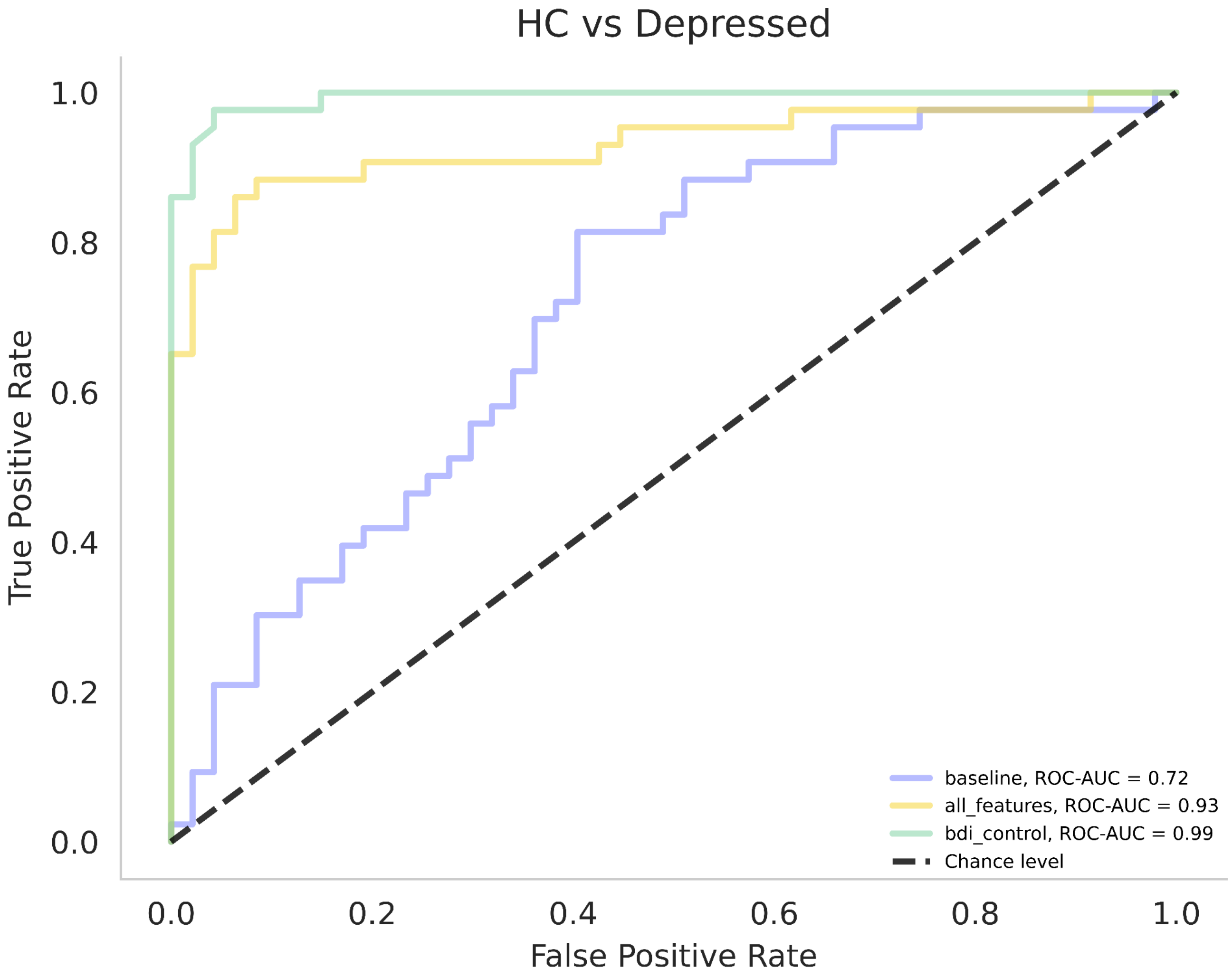

Fig. 2 Receiver operating characteristic curves and areas under the curve (ROC-AUC) for HC vs. Depressed classification models. Legend: HC = healthy controls; baseline = linear model consisting of demographic and clinical data; all_features = SVM speech model; bdi_control = SVM model with BDI-II as the only variable

showing that in MDD patients undergoing sleep deprivation therapy, lower pitch variability was linked to lower depression severity [63]. However, it is worth noting that in this study, the authors assessed within-patient variability while undergoing a specific intervention, whereas the aforementioned studies usually examined variability between participants and treatment as usual.

Interestingly, in our sample, depressed participants spoke significantly louder in both story conditions than did HCs. This is in contrast to most evidence stating that MDD patients on average talk in a lower voice than HCs [28, 31, 64]. We hypothesise various influencing factors that might contribute to these contrasting findings, such as age or task. Our sample was relatively young on average (26±6.5 years). It has been previously demonstrated that symptomatology in MDD varies with age [65, 66], which in turn might also be reflected in speech. Additionally, the paradigms used in studies assessing emotionally charged speech were similar but not identical. For our study presented here, we asked participants to describe a negative and a positive experience they had throughout their lives, whereas for instance Wang and colleagues (2019) asked participants to "please share with us your most wonderful moment and describe it in detail." [31]. For some features, we found differences between the negative and positive stories, which is likely a factor to be considered for the evaluation of speech variables. Additionally, it is worth noting that in our cohort we found that moderately depressed individuals talked in a lower voice than did those with mild depression, although this effect did not reach statistical significance. Alpert and colleagues (2001) reported that depressed individuals talked more loudly than HCs did, although this effect did not reach significance [67].

We found various temporal features associated with the diagnosis of depression, such as utterance duration

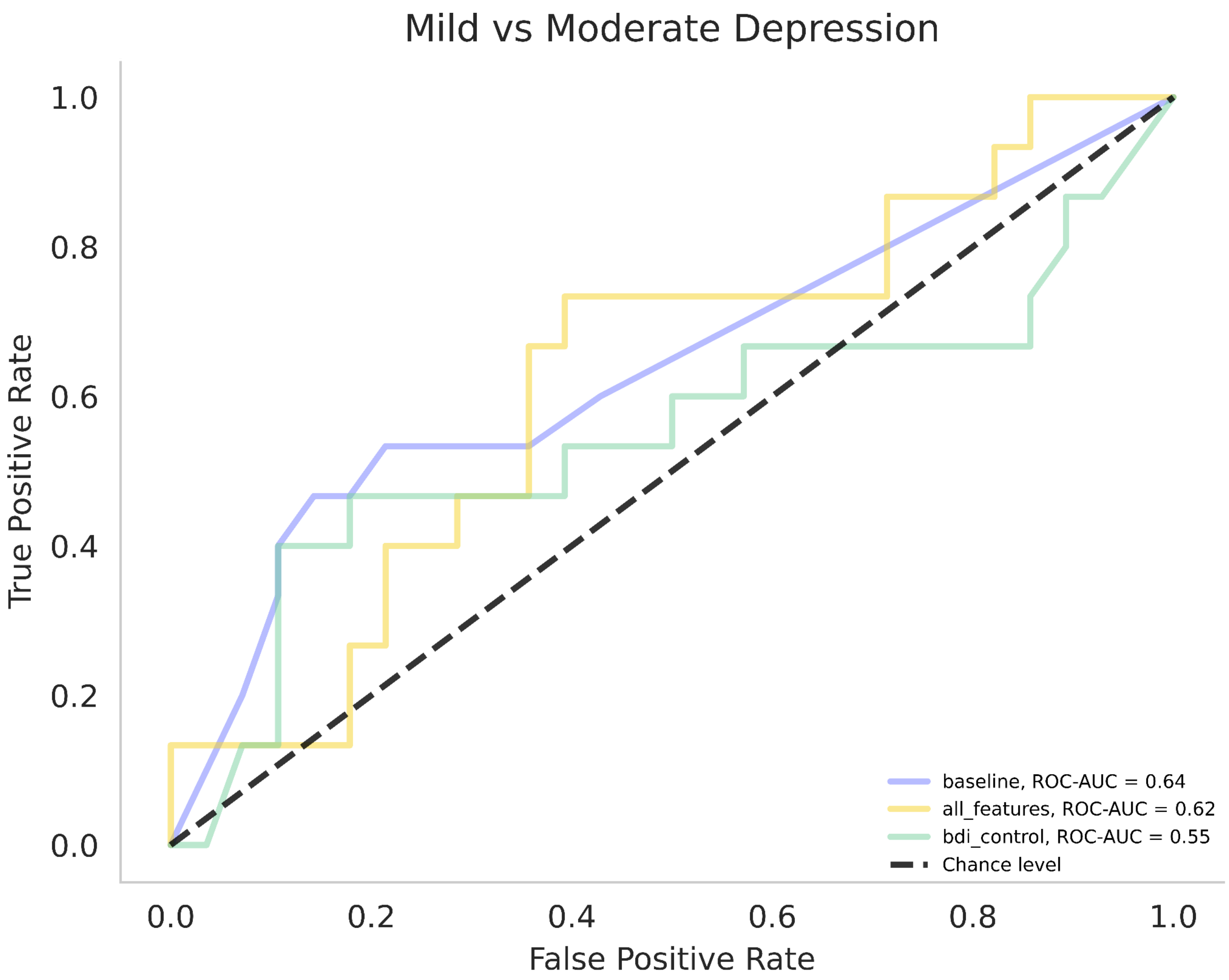

Fig. 3 Receiver operating characteristic curves and areas under the curve (ROC-AUC) for Mild vs. Moderately Depressed classification models. Legend: baseline = Decision Tree Model consisting of demographic and clinical data; all_features = Speech Linear Model; bdi_control = Extra Trees model with BDI-II as the only variable

Table 5 Results of permutation tests to assess the statistical significance of differences between the area under the ROC curves for the HC (Healthy Controls) vs. MDD (major depressive disorder) classification models, and mild vs. moderate MDD patients. Δ = numerical difference

Models	Δ	*p*	Adj. *p*
HC vs. MDD			
BDI-II vs. Speech features	0.06	0.005	**0.01**
BDI-II vs. Baseline	0.28	< 0.001	**< 0.001**
Speech features vs. Baseline	0.21	< 0.001	**< 0.001**
Mild vs. Moderate Depression			
BDI-II vs. Speech features	-0.07	0.66	0.79
BDI-II vs. Baseline	0.09	0.54	0.79
Speech features vs. Baseline	-0.02	0.91	0.91

P = unadjusted *p* value; Adj. *p* = *p* value adjusted for multiple hypothesis testing according to the Benjamini-Hochberg procedure

and pause rate. Our data indicate a lower pause rate for the negative and positive story conditions in depressed individuals than in HCs. This result is seemingly in opposition to the literature describing longer pause times in depressed patients than in HCs [26, 27, 68, 69]. However, our variable "pause rate" is not necessarily comparable to the variables described in the cited publications which focused on the absolute pause duration uncorrected for the total length of speech. In fact, according to our data, group differences (MDD vs. HC) in the absolute variable "pause duration" were not significantly different for either negative or positive storytelling. Additionally, it should be considered that by controlling for the TMT-A, the effect of psychomotor retardation and its effect on speech [70] might have been mitigated. Furthermore, in our data, depressed individuals displayed more pauses than HCs did. This effect, however, did not reach statistical significance in either story condition ($p = 0.11$ and $p = 0.28$, respectively), which is in line with the literature [71, 72].

In our analysis, we noted a difference in word count between MDD patients and HCs. Among HCs, the mean discrepancy in word count between positive and negative narratives was 12 words. However, for those with MDD, the difference was more pronounced, standing at 45 words, with 109 words for the negative story and 64 words for the positive story. These findings align with literature illustrating a tendency for increased use of negative emotion words among individuals experiencing depression compared to HCs [73].

Our speech machine learning model consisting of 10 acoustic features to differentiate between depressed and healthy individuals achieved a performance level that is comparable to the BDI-II model, with an AUC of 0.93 versus 0.99. While our speech model is statistically significantly lower ($p = 0.01$), it still demonstrates strong potential in identifying depression through speech indicators, especially when compared to similar approaches, which yielded lower results (0.66 [74], 0.71 [75], and 0.8 [76]). However, it is worth noting that these studies used different paradigms and samples, thus limiting comparability. Our LM classification based on 50 speech features to differentiate between mildly and moderately depressed patients yielded an AUC of 0.62 compared to the BDI-II XT model with an AUC of 0.55. This numerical superiority, however, did not reach statistical significance ($p = 0.79$). To our knowledge, there is limited literature on machine learning models for differentiating individuals with mild-stage MDD from those with moderate-stage MDD. One study by Shin et al. (2021) investigated the utility of twenty-one voice features to discriminate between minor and major depression [77]. A multilayer processing machine learning method yielded an AUC of 65.9%, similar to our results. However, all of the patients participating in our study suffered from various stages of MDD, in contrast to the differentiation into minor and major depression by Shin and colleagues. Work by Hashim et al. (2017) demonstrated temporal speech features to be predictive of HAMD scores in female patients [78]. For male patients, a combination of temporal and spectral features was predictive of HAMD scores, and temporal features were predictive of BDI-II scores. However, this dimensional approach differed from our categorisation into mild and moderately depressed individuals. Notably, the overlap of features selected for the two classification models that best differentiate between healthy and depressed individuals, as well as between mild and moderately depressed individuals, was minimal, with only two items selected for both models. These results underscore the relevance of distinct speech components in conducting this sort of comparisons. Specifically, the model that differentiates between HC and individuals with MDD consisted solely of acoustic features. In contrast, the model used to distinguish between mild and moderately depressed patients included additional features related to sentiment, word types, and temporal variables.

There are limitations to this study. Since this was an exploratory analysis, no power calculation was performed, which may limit the statistical interpretability and robustness of the findings. No longitudinal data accounting for intraindividual differences are available, which may also allow for a better understanding of treatment response effects on speech. Future research should also focus on associations of speech features with depression on a symptom rather than a syndrome level, i.e., not solely amalgamating symptoms to a numeric score. In addition, our approach to eliciting emotional speech by prompting participants to tell a negative/positive story may limit comparability to other studies where different paradigms were utilised.

Conclusions

Our data show that speech features are associated with depression and that a machine learning model can differentiate between depressed patients and healthy individuals with high accuracy. This performance derived from a two-minute speech recording is comparable to that of an established depression assessment, the BDI-II. Compared to a 10- to 15-minute pen-and-paper assessment, automated speech analysis offers various advantages, such as greater objectivity [79]. Another factor is brevity, which is especially important for trial participants [80]. Furthermore, our approach allows remote access and anonymity without the need for clinician resources to evaluate the questionnaire. In addition, the automated assessment of free speech enables access to qualitative information and may feel less intrusive or clinical, potentially leading to greater engagement and openness from the individual.

Another promising application is remote symptom monitoring, where speech-based markers can be used for continuous, real-time assessment, aiding in early detection of mood changes or depressive relapses. For instance, integrating these markers into smartphone applications could allow patients to be monitored longitudinally outside clinical settings, promoting early intervention. Recent studies, such as Ciampelli et al. (2023), demonstrate that automatic speech recognition (ASR), compared to manual transcribing in combination with semantic natural language processing, effectively analyses speech features without significantly reducing diagnostic accuracy [81]. This technology holds potential for scalable, remote monitoring, enhancing clinical utility and accessibility [82]. Implementing ASR systems for remote depression management could transform treatment by providing clinicians with ongoing, objective assessments, thus improving intervention timeliness and overall patient outcomes.

Future research endeavours in speech analysis among individuals with depression and other psychiatric disorders should consider the development of a standardised speech task protocol. A consortium approach could aid in establishing a common framework. This protocol would ideally incorporate emotionally loaded stimuli while minimising the inclusion of personal information to ensure participant privacy and data protection. Such a standardised protocol would not only facilitate the comparison of results across studies but also streamline the transfer and processing of participant data, ultimately advancing the understanding of voice and speech characteristics in psychiatric disorders. To move towards precision medicine in psychiatry, these findings may aid in the creation of voice-based biomarkers that can enhance the clinical diagnosis and monitoring of psychiatric disorders.

Abbreviations

ASR	Automatic speech recognition
BA	Balanced accuracy
BDI-II	Beck Depression Inventory 2nd edition
BI	Brunet's index
DT	Decision Trees
HAMD	Hamilton Depression Rating Scale
HC	Healthy Controls
LM	Linear Model
LOOCV	Leave-one-out cross-validation
MDD	Major Depressive Disorder
MFCC	Mel-frequency cepstral coefficient
NLP	Natural Language Processing
ROC-AUC	Receiver operating characteristic - area under the curve
STAI	State-Trait-Anxiety Inventory
SVM	Support Vector Machine
TMT	A/B Trail-Making-Test Versions A/B
WMS-R	Wechsler Memory Scale-Revised
XT	Extra Trees

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12888-024-06253-6.

Supplementary Material 1

Acknowledgements
We want to thank all the participants of the study. Moreover, we want to acknowledge the kind support provided by the Brain Imaging Facility of the Interdisciplinary Center for Clinical Research within the Faculty of Medicine at RWTH Aachen University and want to thank Fynn Wagels, Marie Kranz and Lennard Herzberg very much for assisting with the measurements.

Author contributions
FM was responsible for the conceptualisation and original drafting of the manuscript. FD was responsible for data curation, data analysis and for the creation of figures and tables. JS was responsible for data acquisition and contributed to reviewing and writing of the manuscript. JT was responsible for supervision and data analysis. UH was responsible for funding acquisition and supervision. AK was responsible for conceptualisation and supervision and contributed to reviewing and writing of the manuscript. LW was responsible for project administration and data acquisition and contributed to reviewing and writing of the manuscript. All the authors have read and approved the final manuscript.

Funding
Funding for the study was provided by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) 667892/JO 1453/2 – 1. This work was also supported by the FZJ-NST Bilateral Cooperation Program funded by the Forschungszentrum Jülich and the National Research Council of Science & Technology (Global-22-001). Furthermore, the study was partly funded by ki: elements GmbH, Saarbrücken, Germany.

Data availability
Anonymised questionnaire data are available on GitHub (https://github.com/JuliaSchraeder/UnconsciousBias). The speech data generated and analysed during the current study are not publicly available for privacy reasons. The code for speech feature extraction is not publicly available for intellectual property reasons but is available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
The study was approved by the internal ethics committee of the university hospital Aachen, Germany (Ethik-Kommission an der Medizinischen Fakultät der RWTH Aachen; vote number EK 045 – 19). All study procedures were conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants prior to participation in the study.

Consent for publication
Not applicable.

Competing interests
FM, FD, JT and AK are employees of ki: elements GmbH. JS, LW and UH have no competing interests to declare.

Received: 3 July 2024 / Accepted: 4 November 2024
Published online: 12 November 2024

References

1. de la Torre JA, Vilagut G, Ronaldson A, Serrano-Blanco A, Martín V, Peters M, et al. Prevalence and variability of current depressive disorder in 27 European countries: a population-based study. Lancet Public Health. 2021;6(10):e729–38.
2. Greenberg PE, Fournier AA, Sisitsky T, Simes M, Berman R, Koenigsberg SH, et al. The economic burden of adults with major depressive disorder in the United States (2010 and 2018). PharmacoEconomics. 2021;39(6):653–65.
3. IsHak WW, Mirocha J, James D, Tobia G, Vilhauer J, Fakhry H, et al. Quality of life in major depressive disorder before/after multiple steps of treatment and one-year follow-up. Acta Psychiatr Scand. 2015;131(1):51–60.
4. Iancu SC, Wong YM, Rhebergen D, van Balkom AJLM, Batelaan NM. Long-term disability in major depressive disorder: a 6-year follow-up study. Psychol Med. 2020;50(10):1644–52.
5. Marx W, Penninx BWJH, Solmi M, Furukawa TA, Firth J, Carvalho AF, et al. Major depressive disorder. Nat Rev Dis Primer. 2023;9(1):1–21.
6. American Psychiatric Association. DSM–5 Task Force. Diagnostic and statistical manual of mental disorders (DSM–5®). American Psychiatric Association; 2013. p. 947.
7. Marwaha S, Palmer E, Suppes T, Cons E, Young AH, Upthegrove R. Novel and emerging treatments for major depression. Lancet. 2023;401(10371):141–53.

8. Hansson O, Blennow K, Zetterberg H, Dage J. Blood biomarkers for Alzheimer's disease in clinical practice and trials. Nat Aging. 2023;3(5):506–19.
9. Lee Y, Ragguett RM, Mansur RB, Boutilier JJ, Rosenblat JD, Trevizol A, et al. Applications of machine learning algorithms to predict therapeutic outcomes in depression: a meta-analysis and systematic review. J Affect Disord. 2018;241:519–32.
10. Jacobson NC, Weingarden H, Wilhelm S. Digital biomarkers of mood disorders and symptom change. Npj Digit Med. 2019;2(1):1–3.
11. Schultebraucks K, Yadav V, Galatzer-Levy IR. Utilization of Machine Learning-Based Computer Vision and Voice Analysis To Derive Digital Biomarkers of Cognitive Functioning in Trauma survivors. Digit Biomark. 2020;16–23.
12. Malgaroli M, Schultebraucks K. Artificial intelligence and posttraumatic stress disorder (PTSD): an overview of advances in research and emerging clinical applications. Eur Psychol. 2020;25(4):272–82.
13. Kappen M, Vanderhasselt MA, Slavich GM. Speech as a promising biosignal in precision psychiatry. Neurosci Biobehav Rev. 2023;148:105121.
14. de Boer JN, Voppel AE, Brederoo SG, Schnack HG, Truong KP, Wijnen FNK, et al. Acoustic speech markers for schizophrenia-spectrum disorders: a diagnostic and symptom-recognition tool. Psychol Med. 2023;53(4):1302–12.
15. Koops S, Brederoo SG, de Boer JN, Nadema FG, Voppel AE, Sommer IE. Speech as a Biomarker for Depression. CNS Neurol Disord Drug Targets. 2023;22(2):152–60.
16. Marmar CR, Brown AD, Qian M, Laska E, Siegel C, Li M, et al. Speech-based markers for posttraumatic stress disorder in US veterans. Depress Anxiety. 2019;36(7):607–16.
17. Eyben F, Scherer KR, Schuller BW, Sundberg J, Andre E, Busso C, et al. The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing. IEEE Trans Affect Comput. 2016;7(2):190–202.
18. König A, Tröger J, Mallick E, Mina M, Linz N, Wagnon C, et al. Detecting subtle signs of depression with automated speech analysis in a non-clinical sample. BMC Psychiatry. 2022;22(1):830.
19. König A, Mina M, Schäfer S, Linz N, Tröger J. Predicting Depression Severity from spontaneous Speech as prompted by a virtual Agent. Eur Psychiatry. 2023;66(S1):S157–8.
20. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. Speech Commun. 2015;71:10–49.
21. Gupta R, Malandrakis N, Xiao B, Guha T, Van Segbroeck M, Black M et al. Multimodal Prediction of Affective Dimensions and Depression in Human-Computer Interactions. In: Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge. Orlando Florida USA: ACM; 2014 [cited 2024 Jun 5];33–40. https://doi.org/10.1145/2661806.2661810
22. Ettore E, Müller P, Hinze J, Riemenschneider M, Benoit M, Giordana B, et al. Digital phenotyping for Differential diagnosis of major depressive episode: Narrative Review. JMIR Ment Health. 2023;10(1):e37225.
23. Low DM, Bentley KH, Ghosh SS. Automated assessment of psychiatric disorders using speech: a systematic review. Laryngoscope Investig Otolaryngol. 2020;5(1):96–116.
24. Horwitz R, Quatieri TF, Helfer BS, Yu B, Williamson JR, Mundt J. On the relative importance of vocal source, system, and prosody in human depression. 2013 IEEE Int Conf Body Sens Netw. 2013;1–6.
25. Kiss G, Vicsi K. Mono- and multi-lingual depression prediction based on speech processing. Int J Speech Technol. 2017;20(4):919–35.
26. Mundt JC, Vogel AP, Feltner DE, Lenderking WR. Vocal acoustic biomarkers of depression severity and treatment response. Biol Psychiatry. 2012;72(7):580–7.
27. Yamamoto M, Takamiya A, Sawada K, Yoshimura M, Kitazawa M, Liang K, ching et al. Using speech recognition technology to investigate the association between timing-related speech features and depression severity. Hashimoto K, editor. PLOS ONE. 2020;15(9):e0238726.
28. Alghowinem S, Goecke R, Wagner M, Epps J, Breakspear M, Parker G. Detecting depression: A comparison between spontaneous and read speech. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. ieeexplore.ieee.org; 2013;7547–51.
29. Cummins N, Sethu V, Epps J, Schnieder S, Krajewski J. Analysis of acoustic space variability in speech affected by depression. Speech Commun. 2015;75:27–49.
30. Taguchi T, Tachikawa H, Nemoto K, Suzuki M, Nagano T, Tachibana R, et al. Major depressive disorder discrimination using vocal acoustic features. J Affect Disord. 2018;225:214–20.
31. Wang J, Zhang L, Liu T, Pan W, Hu B, Zhu T. Acoustic differences between healthy and depressed people: a cross-situation study. BMC Psychiatry. 2019;19(1):300.
32. Trifu R, Nemes B, Bodea-Hațegan C, Cozman D. Linguistic indicators of language in major depressive disorder (MDD). An evidence based research. J Evid-Based Psychother. 2017;17:105–28.
33. Arevian AC, Bone D, Malandrakis N, Martinez VR, Wells KB, Miklowitz DJ et al. Clinical state tracking in serious mental illness through computational analysis of speech. Scilingo EP, editor. PLOS ONE. 2020;15(1):e0225695.
34. Shinohara S, Nakamura M, Omiya Y, Higuchi M, Hagiwara N, Mitsuyoshi S, et al. Depressive Mood Assessment Method based on emotion level derived from Voice: comparison of Voice Features of Individuals with Major Depressive Disorders and Healthy Controls. Int J Environ Res Public Health. 2021;18(10):5435.
35. Stasak B, Epps J, Cummins N, Goecke R. An Investigation of Emotional Speech in Depression Classification. In: Interspeech 2016. ISCA; 2016 [cited 2024 Oct 16];485–9. https://www.isca-archive.org/interspeech_2016/stasak16_interspeech.html
36. Aharonson V, de Nooy A, Bulkin S, Sessel G. Automated Classification of Depression Severity Using Speech - A Comparison of Two Machine Learning Architectures. In: 2020 IEEE International Conference on Healthcare Informatics (ICHI). 2020 [cited 2024 Oct 16];1–4. https://ieeexplore.ieee.org/document/9374335
37. Shinohara S, Toda H, Nakamura M, Omiya Y, Higuchi M, Takano T, et al. Evaluation of the severity of Major Depression using a Voice Index for Emotional Arousal. Sensors. 2020;20(18):5041.
38. Kwon N, Kim S. Depression Severity Detection Using Read Speech with a Divide-and-Conquer Approach. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2021 [cited 2024 Oct 16];633–7. https://ieeexplore.ieee.org/document/9629868
39. Schräder J, Herzberg L, Jo HG, Hernandez-Pena L, Koch J, Habel U et al. Neurophysiological pathways of unconscious emotion Processing in Depression: insights from a simultaneous electroencephalography-functional magnetic resonance imaging measurement. Biol Psychiatry Cogn Neurosci Neuroimaging. 2024;S2451–9022(24)00193–9.
40. Beesdo-Baum K, Zaudig M, Wittchen HU. SCID–5-CV Strukturiertes Klinisches Interview für DSM–5-Störungen–Klinische Version: Deutsche Bearbeitung des Structured Clinical Interview for DSM–5 Disorders–Clinician Version von Michael B. First, Janet BW Williams, Rhonda S. Karg, Robert L. Hogrefe; 2019 [cited 2024 Apr 22]. https://www.testzentrale.de/shop/strukturiertes-klinisches-interview-fuer-dsm–5r-stoerungen-klinische-version.html
41. Beck AT, Steer RA, Brown G. Beck Depression Inventory–II (BDI-II). 1996 [cited 2024 May 13]. https://doi.org/10.1037/t00742-000
42. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry. 1960;23:56–61.
43. Reitan RM. The relation of the trail making test to organic brain damage. J Consult Psychol. 1955;19(5):393–4.
44. Härting C. Wechsler-Gedächtnistest - Revidierte Fassung: WMS-R; manual ; deutsche Adaptation Der Revidierten Fassung Der Wechsler Memory scale. Huber; 2000;125.
45. Spielberger C, Gorsuch R, Lushene R, Vagg P, Jacobs G. Manual for the state-trait anxiety inventory (form Y1 – Y2). Palo Alto, CA: Consulting Psychologists Press; 1983;IV.
46. König A, Mallick E, Tröger J, Linz N, Zeghari R, Manera V, et al. Measuring neuropsychiatric symptoms in patients with early cognitive decline using speech analysis. Eur Psychiatry. 2021;64(1):e64.
47. König A, Linz N, Zeghari R, Klinge X, Tröger J, Alexandersson J, et al. Detecting apathy in older adults with cognitive disorders using automatic speech analysis. J Alzheimers Dis. 2019;69(4):1183–93.
48. Cummins N, Dineley J, Conde P, Matcham F, Siddi S, Lamers F, et al. Multilingual markers of depression in remotely collected speech samples: a preliminary analysis. J Affect Disord. 2023;341:128–36.
49. Ladefoged P. Elements of acoustic phonetics. 2nd ed. Chicago: University of Chicago Press; 1996. p. 216.
50. Ladefoged P, Johnson K. A course in Phonetics. 6th Edition. Boston, MA, USA: Michael Rosenberg; 2011.
51. Teixeira JP, Oliveira C, Lopes C. Vocal acoustic analysis – jitter, Shimmer and HNR parameters. Procedia Technol. 2013;9:1112–22.
52. Nakagawa S, Asakawa K, Wang L. Speaker recognition by combining MFCC and phase information. In: Interspeech 2007. ISCA; 2007 [cited 2024 May 16];2005–8. https://www.isca-archive.org/interspeech_2007/nakagawa07_interspeech.html

53. Zellner B. Pauses and the temporal structure of Speech. Fundamentals of speech synthesis and speech recognition. Chichester: John Wiley; 1994. pp. 41–62.
54. Lindsay H, Tröger J, König A. Language Impairment in Alzheimer's Disease—Robust and Explainable Evidence for AD-Related Deterioration of Spontaneous Speech Through Multilingual Machine Learning. Front Aging Neurosci. 2021 May 19 [cited 2024 Apr 22];13. https://www.frontiersin.org/articles/https://doi.org/10.3389/fnagi.2021.642033
55. Qi P, Zhang Y, Zhang Y, Bolton J, Manning CD. Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. Online: Association for Computational Linguistics; 2020 [cited 2023 Oct 26];101–8. https://aclanthology.org/2020.acl-demos.14
56. Socher R, Perelygin A, Wu J, Chuang J, Manning CD, Ng A et al. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In: Yarowsky D, Baldwin T, Korhonen A, Livescu K, Bethard S, editors. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Seattle, Washington, USA: Association for Computational Linguistics; 2013 [cited 2024 Jun 26];1631–42. Available from: https://aclanthology.org/D13-1170.
57. Buyukdura JS, McClintock SM, Croarkin PE. Psychomotor retardation in depression: Biological underpinnings, measurement, and treatment. Prog Neuropsychopharmacol Biol Psychiatry. 2011;35(2):395–409.
58. Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. Phys Rev E. 2004;69(6):066138.
59. Benjamini Y, Hochberg Y. Controlling the false Discovery rate: a practical and powerful Approach to multiple testing. J R Stat Soc Ser B Methodol. 1995;57(1):289–300.
60. Sundberg J, Nordenberg M. Effects of vocal loudness variation on spectrum balance as reflected by the alpha measure of long-term-average spectra of speech. J Acoust Soc Am. 2006;120(1):453–7.
61. Brunet E. Le Vocabulaire de Jean Giraudoux. In: Structure Et Evolution. Geneve; 1978 [cited 2024 May 16]. Available from: https://books.google.de/books?hl=en&lr=&id=t1COajFe6l0C&oi=fnd&pg=PA1&ots=ttkEfKdxvL&sig=RjTZhHnw-gBG8xYzQm97mugD1W8&redir_esc=y#v=onepage&q&f=false
62. Sanchez MH, Vergyri D, Ferrer L, Richey C, Garcia P, Knoth B et al. Using prosodic and spectral features in detecting depression in elderly males. In: Twelfth Annual Conference of the International Speech Communication Association. isca-speech.org; 2011.
63. Wadle LM, Ebner-Priemer UW, Foo JC, Yamamoto Y, Streit F, Witt SH, et al. Speech features as predictors of momentary depression severity in patients with depressive disorder undergoing sleep deprivation therapy: ambulatory Assessment Pilot Study. JMIR Ment Health. 2024;11:e49222.
64. Wang Y, Liang L, Zhang Z, Xu X, Liu R, Fang H et al. Fast and accurate assessment of depression based on voice acoustic features: a cross-sectional and longitudinal study. Front Psychiatry. 2023 [cited 2023 Jun 21];14. https://www.frontiersin.org/articles/https://doi.org/10.3389/fpsyt.2023.1195276
65. Hybels CF, Landerman LR, Blazer DG. Age differences in symptom expression in patients with major depression. Int J Geriatr Psychiatry. 2012;27(6):601–11.
66. Wagner S, Wollschläger D, Dreimüller N, Engelmann J, Herzog DP, Roll SC, et al. Effects of age on depressive symptomatology and response to antidepressant treatment in patients with major depressive disorder aged 18 to 65 years. Compr Psychiatry. 2020;99:152170.
67. Alpert M, Pouget ER, Silva RR. Reflections of depression in acoustic measures of the patient's speech. J Affect Disord. 2001;66(1):59–69.
68. Esposito A, Esposito AM, Likforman-Sulem L, Maldonato MN, Vinciarelli A et al. On the Significance of Speech Pauses in Depressive Disorders: Results on Read and Spontaneous Narratives. In: Esposito A, Faundez-Zanuy M, Esposito AM, Cordasco G, Drugman T, Solé-Casals J, editors. Recent Advances in Nonlinear Speech Processing. Cham: Springer International Publishing; 2016 [cited 2024 May 29];73–82. https://doi.org/10.1007/978-3-319-28109-4_8
69. Liu Z, Kang H, Feng L, Zhang L. Speech pause time: A potential biomarker for depression detection. In: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2017 [cited 2024 May 29];2020–5. https://ieeexplore.ieee.org/abstract/document/8217971
70. Bennabi D, Vandel P, Papaxanthis C, Pozzo T, Haffen E. Psychomotor retardation in Depression: a systematic review of Diagnostic, Pathophysiologic, and therapeutic implications. BioMed Res Int. 2013;2013:158746.
71. Mundt JC, Snyder PJ, Cannizzaro MS, Chappie K, Geralts DS. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. J Neurolinguistics. 2007;20(1):50–64.
72. Wolters MK, Ferrini L, Farrow E, Tatar AS, Burton CD. Tracking depressed mood using speech pause patterns. In internationalphoneticassociation.org; 2015. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0811.pdf
73. Tølbøll KB. Linguistic features in depression: a meta-analysis. J Lang Works - Sprogvidenskabeligt Stud. 2019;4(2):39–59.
74. Bauer JF, Gerczuk M, Schindler-Gmelch L, Amiriparian S, Ebert DD, Krajewski J, et al. Validation of machine learning-based Assessment of Major Depressive Disorder from Paralinguistic Speech characteristics in Routine Care. Depress Anxiety. 2024;2024(1):9667377.
75. Hansen L, Zhang YP, Wolf D, Sechidis K, Ladegaard N, Fusaroli R. A generalizable speech emotion recognition model reveals depression and remission. Acta Psychiatr Scand. 2022;145(2):186–99.
76. Di Y, Wang J, Li W, Zhu T. Using i-vectors from voice features to identify major depressive disorder. J Affect Disord. 2021;288:161–6.
77. Shin D, Cho WI, Park CHK, Rhee SJ, Kim MJ, Lee H, et al. Detection of minor and major depression through Voice as a Biomarker using machine learning. J Clin Med. 2021;10(14):3046.
78. Hashim NW, Wilkes M, Salomon R, Meggs J, France DJ. Evaluation of Voice Acoustics as predictors of Clinical Depression scores. J Voice off J Voice Found. 2017;31(2):e2561–6.
79. Fagherazzi G, Fischer A, Ismael M, Despotovic V. Voice for health: the use of vocal biomarkers from research to clinical practice. Digit Biomark. 2021;5(1):78–88.
80. ePRO Consortium, Bodart S, Byrom B, Crescioni M, Eremenco S, Flood E. Perceived Burden of Completion of patient-reported outcome measures in clinical trials: results of a preliminary study. Ther Innov Regul Sci. 2019;53(3):318–23.
81. Ciampelli S, Voppel AE, de Boer JN, Koops S, Sommer IEC. Combining automatic speech recognition with semantic natural language processing in schizophrenia. Psychiatry Res. 2023;325:115252.
82. Ramanarayanan V. Multimodal Technologies for Remote Assessment of Neurological and Mental Health. J Speech Lang Hear Res JSLHR. 2024;1–13.

Publisher's note