ARTICLE OPEN

Check for updates

Objective speech measures capture depressive symptoms and associated cognitive difficulties

Micaela Wiseman[1,2], Rachel Yep[1], Madeline Wood Alexander[1,3], Christopher B. Pople[1,2], Lucas Perri[1], Georgia Gopinath[1], Maria Vasileiadi[1], Jessica Robin[4], Michael J. Spilka[4], William Simpson[4], Yana Yunusova[1,3], Douglas P. Munoz[5], Brian C. Coe[5], Donald Brien[5], Sean Nestor[1,6], Nir Lipsman[1,2,6,7,8], Peter Giacobbe[1,2,6] and Jennifer S. Rabin[1,2,3,8]✉

Psychiatry lacks objective biomarkers for assessing depression, relying instead on subjective measures, such as the Hamilton Depression Rating Scale (HAMD-17). This study examined whether speech features could serve as objective markers of depressive symptoms and its associated cognitive difficulties. Sixty-six individuals with major depressive disorder (MDD) and 54 non-depressed control participants completed a speech assessment, responding to the prompt: *"Please tell me how you are feeling today."* Linguistic (valence, emotional intensity, agency) and acoustic (pitch, pitch variance, speech rate, time spent pausing) features were derived from natural language processing. These speech features were analyzed individually and collectively as a composite score representing overall speech disturbance. A subset of participants (40 with MDD, 38 controls) also completed a validated executive function task. ANCOVA models compared speech features between groups. Linear regression models examined associations between speech features, depression severity (HAMD-17), and performance on an executive function task. Compared to controls, individuals with MDD used language that was more negatively valenced, emotionally intense, and less agentic. They also demonstrated lower pitch, slower speech rate, and more time spent pausing. The composite speech score also differed between groups. Speech features and executive function were not associated with depression severity, as measured by the HAMD-17. However, several speech features were associated with executive function. Taken together, these findings suggest that speech features may provide a scalable, objective method for detecting depressive symptoms and associated executive difficulties.

Translational Psychiatry (2025)15:525 ; https://doi.org/10.1038/s41398-025-03728-2

INTRODUCTION

Depression is a prevalent and disabling mental health condition, characterized by a range of symptoms including low mood, anhedonia, somatic changes, and cognitive difficulties [1]. Unlike other fields of medicine, psychiatry currently lacks validated objective biomarkers for assessing depression. Instead, symptoms are typically measured using self- and clinician-reported measures, with the Hamilton Depression Rating Scale (HAMD-17) being among the most widely used. However, these traditional assessments can be influenced by recall bias, fluctuations in mood state, individuals' insight into their symptoms, and clinician subjectivity [2, 3]. Additionally, such measures often fail to fully capture cognitive symptoms, which are both common in depression and strongly associated with poor functioning and quality of life [4].

Recent advancements in artificial intelligence (AI) and natural language processing (NLP) have sparked interest in speech as a potential biomarker for depression [5, 6]. Speech-based biomarkers offer several advantages: they are scalable, non-invasive, and can be collected remotely, making them an appealing complement to traditional assessment scales. Speech output reflects linguistic content (what is said) and acoustic qualities (how it is said), both of which are altered in depression [7, 8]. In terms of linguistic features, studies analyzing social media posts, psychotherapy transcripts, essays, and free speech show that individuals with depression tend to use language that is more negatively valenced [9–15] (e.g., "sad", "failure", vs. "happy", "comfort"), more emotionally intense (e.g., "despair", "panic" vs. "fatigue", "relaxed"), and reflective of lower agency (e.g., "powerless", "hopeless" vs. "confidence", "capable") [16–18]. Studies also show that more negatively valenced language is associated with greater depression severity [10, 19, 20].

With respect to acoustic features, clinicians have long observed that individuals with depression speak in a monotonic, slow, and effortful manner, observations that have now been systematically documented. For example, studies consistently report that individuals with depression exhibit lower pitch [21–28], less pitch variability [24, 27], slower speech rate [25, 29–35] and longer pauses [25, 34, 36] compared to non-depressed participants. Additionally, several studies have reported that greater depression severity is associated with lower and less variable pitch [37–40], slower speech rate [29, 37, 39, 41–43], and longer pauses [39, 42, 43].

[1]Sunnybrook Research Institute, Toronto, Canada. [2]Institute of Medical Sciences, University of Toronto, Toronto, Canada. [3]Rehabilitation Sciences Institute, University of Toronto, Toronto, Canada. [4]Cambridge Cognition, Cambridge, United Kingdom. [5]Centre for Neuroscience Studies, Queens University, Kingston, Canada. [6]Department of Psychiatry, University of Toronto, Toronto, Canada. [7]Department of Surgery, University of Toronto, Toronto, Canada. [8]Department of Medicine (Neurology), University of Toronto, Toronto, Canada. ✉email: jennifer.rabin@sri.utoronto.ca

Received: 29 May 2025 Revised: 29 September 2025 Accepted: 21 October 2025
Published online: 17 November 2025

Speech features may also capture cognitive difficulties. For example, a study of community-dwelling participants in the Framingham Heart Study found that lower and less variable pitch, along with shorter speech segment lengths, were associated with cognitive impairment [44]. Similarly, individuals with mild cognitive impairment have been shown to exhibit slower speaking rates and longer pauses compared to healthy controls [45]. However, the potential of speech features to capture cognitive difficulties in individuals with depression has not yet been examined.

Changes in cognitive function is a recognized criterion of depression, with difficulties in executive function well-documented among individuals with the disorder [46]. Cognitive models of depression highlight deficits in executive function as a core mechanism underlying symptoms, such as persistent rumination, reduced goal-directed behavior, and difficulties disengaging from negative information [47–49]. However, traditional depression measures, such as the HAMD-17, often fail to fully capture cognitive symptoms [4, 50]. This gap is significant, as cognitive difficulties in depression are closely linked to poor functional outcomes and reduced quality of life [4, 51]. These limitations highlight the need for more comprehensive assessment tools that capture the full range of depression symptoms, including cognitive symptoms. To our knowledge, no prior studies have examined the association between speech features and cognitive performance in individuals with depression. Investigating this link may lay the groundwork for developing more objective approaches to depression assessment.

The present study aimed to address this gap by examining whether speech features could serve as an objective indicator of depression and its associated cognitive difficulties, with a particular focus on executive function. Study participants included individuals with a diagnosis of major depressive disorder (MDD) and non-depressed controls, who completed a brief speech assessment and the interleaved pro/anti saccade task (IPAST), a validated and widely used measure of executive function [52].

The study aimed to address three main questions: 1) Do speech features differ between individuals with MDD and non-depressed control participants? 2) Are speech features associated with depression severity, as measured by the HAMD-17? and 3) Are speech features associated with executive function, as measured by the IPAST? Based on previous research, we anticipated that speech features would differ between the two groups and would be associated with both depression severity and executive function performance.

MATERIALS AND METHODS

Participants

Participants with MDD were recruited from the Harquail Centre for Neuromodulation at Sunnybrook Health Sciences Centre. All participants had treatment-resistant MDD, which was defined as a failure to respond to a minimum of two adequately dosed antidepressant medications and a HAMD-17 score of ≥16 [53]. After clinical screening but prior to receiving any neuromodulation intervention (e.g., repetitive transcranial magnetic stimulation, deep brain stimulation, or high-frequency focused ultrasound), participants were enrolled and tested as part of the present study.

Inclusion criteria for the present study were as follows: (1) a diagnosis of treatment-resistant unipolar depression with a current major depressive episode, confirmed by a psychiatrist in accordance with DSM-5 criteria; (2) age 18 years or older; (3) completion of at least 8 years of education; and (4) fluency in English. Exclusion criteria included: (1) a history of substance use disorder with active use in the past 6 months; (2) a diagnosis of a neurological or neurocognitive disorder; and (3) significant comorbid psychiatric or medical conditions (e.g., bipolar disorder, schizophrenia, uncontrolled diabetes).

Non-depressed control participants were recruited from the community through online advertisements as well as posters placed in hospitals and public community spaces. Eligibility criteria included: (1) age 18 years or older; (2) completion of at least 8 years of education; and (3) fluency in English. Exclusion criteria were consistent with those applied to participants with depression, with the additional requirement that controls had no current or past history of psychiatric illness or related treatment (e.g., psychiatric medication, psychotherapy, or hospitalization due to a psychiatric illness).

All procedures were reviewed and approved by the Sunnybrook Ethics Review Board (Project Identification Number: 2106). Participants provided informed consent before undergoing study procedures.

Speech assessment

Winterlight Labs' speech assessment tool [54] was used to collect speech samples via an iOS/Android application. Participants responded to the prompt, *"Please tell me how you are feeling today."* Participants could speak for as long as they wished. Assessments were completed either at Sunnybrook Health Sciences Centre or remotely in a quiet room.

Speech recordings were analyzed using Winterlight Labs' validated software, which employs NLP to extract acoustic and linguistic features for statistical analysis. The software and processing pipeline have been described elsewhere [55]. For this study, we focused on linguistic and acoustic speech features previously associated with depression [5, 6, 9]. Linguistic features included valence, emotional intensity, and agency of the language used, whereas acoustic features included pitch, pitch variance, speech rate, and time spent pausing (summarized in Table 1). We did not examine features sensitive to phoneme variation, such as jitter, shimmer, or harmonics-to-noise ratio, as our task involved spontaneous speech, which produces greater variability in phoneme articulation than controlled, read-speech tasks [56, 57]. Similarly, we did not include mel-frequency cepstral coefficients, as they can be affected by inconsistent recording environments [58, 59].

Executive function

Executive function was assessed using the IPAST, an eye-tracking task that provides a sensitive and robust measure of executive function [60, 61]. We selected the IPAST over standard neuropsychological tests because it is highly sensitive to core executive processes [62–64], such as attentional control, inhibition, and cognitive flexibility, all of which have been implicated in depression. Moreover, because the IPAST relies on eye movements rather than verbal or manual responses, it reduces potential confounding effects associated with motor output [65].

Participants completed the IPAST in a dark room with their heads supported in a headrest. They were positioned 60 cm from a 17-inch LCD monitor (1280 × 1024 pixel resolution, 32-bit color, 60 Hz refresh rate). Eye position was tracked using an infrared video-based eye tracker (Eyelink 1000 Plus, SR Research Ltd, Toronto, ON, Canada) at a sampling rate of 500 Hz. Before starting the task, each participant underwent a 9-point array calibration and validation procedure, followed by a practice session.

In the IPAST, participants completed a series of trials in which they were instructed to either look toward a visual stimulus (pro-saccade) or suppress this response and look in the opposite direction (anti-saccade). Pro- and anti-saccade trials were intermixed throughout the task. Pro-saccade reaction time (RT) reflects processing speed, whereas anti-saccade RT reflects executive function. The latter was the primary variable of interest, with slower RTs indicating greater difficulty inhibiting the automatic response to look toward a visual stimulus, and therefore poorer executive function [52, 60, 61]. Participants completed 240 trials of the IPAST. The full task protocol is described in Yep et al. [52]. All eye tracking data were collected at Sunnybrook Health Sciences Centre. IPAST data were processed using previously described methods [66].

Statistical analysis

Analyses were conducted in R version 4.3.3. Separate models were performed for each speech feature of interest: valence, emotional intensity, agency (linguistic features), as well as mean pitch, pitch variance, speech rate, and time spent pausing (acoustic features). A composite speech score was calculated by converting each raw speech feature to a z-score based on sample means and standard deviations. For features negatively associated with depression, z-scores were sign-reversed to ensure consistent directionality. The adjusted z-scores were then averaged to produce a single composite measure, with higher scores reflecting greater speech alteration.

Demographic variables, including age, sex, and years of education, were compared between individuals with MDD and non-depressed control participants using student's *t*-tests for continuous variables and chi-squared tests for categorical variables.

Table 1. Linguistic and acoustic speech features assessed and their operational definitions.

Feature	Type	Description	Unit of Measurement
Valence	Linguistic	Measures the emotional positivity or negativity of words. Scores range from 1–9 where lower scores indicate the use of more negatively valenced language (e.g., "murder" = 1.48 vs. "happiness" = 8.48).	Score based on standardized norms from psycholinguistic research [77].
Emotional intensity	Linguistic	Measures the level of emotional intensity in words. Scores range from 1–9 where higher scores correspond to more emotionally charged language (e.g., "insanity" = 7.79 vs. "calm" = 1.67).	Score based on standardized norms from psycholinguistic research [77].
Agency	Linguistic	Measures the degree of control or agency conveyed by words. Scores range from 1–9 where lower scores indicate language associated with lower control or agency (e.g., "uncontrollable" = 2.18 vs. "successful" = 7.71).	Score based on standardized norms from psycholinguistic research [77].
Mean pitch	Acoustic	Measures the average rate at which the vocal folds vibrate during speech. This vibration rate determines perceived voice pitch. Higher values correspond to higher-pitched voices and lower values correspond to lower-pitched voices.	Hertz (Hz)
Pitch variance	Acoustic	Measures fluctuations in pitch and provides a measure of prosody. Higher values represent more prosodic speech while lower values represent more monotone or steady speech.	Hertz (Hz)
Speech rate	Acoustic	Measures the speed of speech, with lower scores reflecting slower speech.	Words per minute
Pauses	Acoustic	Measures the proportion of the audio recording consisting of pauses lasting 1–2 s, controlling for the total length of the recording. Higher scores reflect a greater proportion of time spent pausing.	Proportion of time spent pausing

To address the first question—whether speech features differed between individuals with MDD and non-depressed controls—we conducted a series of one-way ANCOVA models. Group (MDD vs. control) served as the independent variable, and each speech feature/the speech composite served as the dependent variable in separate models. All models included the following covariates: age (continuous), sex (categorical: male or female), years of education (continuous), and the duration of the audio recording (continuous). In sensitivity analyses, we additionally adjusted for testing location (in-person vs. remote) and age at English acquisition. These variables were not included in the primary models because all participants were required to be fluent in English, and only a small subset of participants completed the task in person.

Additionally, a Receiver Operating Characteristic (ROC) analysis was performed to assess the classification accuracy of each speech feature and the speech composite, while adjusting for age, sex, years of education and audio recording duration. While ANCOVA models tested for group differences in mean values, ROC analysis provided a complementary measure of how well each feature distinguished between participants with and without MDD, highlighting their relevance for speech-based screening of depression.

To address the second question—whether speech features were associated with depression severity, as measured by the HAMD-17—we used linear regression models. In these analyses, speech features served as the independent variables, and HAMD-17 scores as the dependent variable. Separate models were performed for each speech feature/the speech composite. Covariates were identical to those used in the ANCOVA models.

To address the third question —whether speech features were associated with executive function, as measured by the IPAST—we used linear regression models. In these analyses, individual speech features/the speech composite served as the predictor variable, and anti-saccade RTs served as the outcome variable. As above, models were adjusted for age, sex, years of education, and audio recording duration. Because the executive measure from the IPAST is based on RT, we included pro-saccade RT as a covariate to control for general processing speed. This adjustment accounted for individual differences in processing speed and allowed us to more accurately capture the executive component of the task.

To further assess whether these associations were specific to executive function, we performed additional linear regression analyses to examine whether speech features were associated with pro-saccade RT. In these models, individual speech features/the speech composite served as the predictor variable, and pro-saccade RT served as the outcome variable. As above, models adjusted for age, sex, years of education, and audio recording duration.

Statistical significance was evaluated using a two-tailed alpha level of 0.05. To account for multiple comparisons, we applied a False Discovery Rate (FDR) correction using the Benjamini-Hochberg procedure, separately for each of the three research questions. Both unadjusted and FDR-corrected *p*-values are reported. Effect sizes for the ANCOVAs are reported as partial η^2, and effect sizes for the linear regression models are reported as Cohen's f^2. No a priori power analysis was performed. A sensitivity analysis indicated that, with N = 120 and four covariates ($df_1 = 1$, $df_2 = 114$; $\alpha = 0.05$), 1-df effects (group contrasts and single-predictor slopes) had ~80% power to detect partial $r \approx 0.25$ (partial $\eta^2 \approx 0.065$; $f^2 \approx 0.07$); for the anti-saccade models (N = 78, five covariates; $df_1 = 1$, $df_2 = 71$), ~80% power corresponded to partial $r \approx 0.32$ (partial $\eta^2 \approx 0.10$; $f^2 \approx 0.11$).

RESULTS

Participant characteristics

Table 2 summarizes the demographic and clinical characteristics of study participants. A total of 66 individuals with MDD and 54 non-depressed control participants completed the speech assessment. A subset of these participants also completed the IPAST eye-tracking task. This included 40 individuals with MDD (mean age = 45.0 ± 15.6, 55% female, HAMD-17 = 21.0 ± 4.3) and 38 control participants (mean age = 41.2 ± 15.7, 81.6% female).

Prior to the main analyses, we examined associations among the seven speech features across the full sample using partial Pearson correlations, adjusting for age, sex, years of education, and audio recording duration. After applying FDR correct for multiple comparisons, several significant associations emerged (Figure S1).

Question 1: Do speech features differ between individuals with MDD and non-depressed control participants?

As summarized in Table 3 and Fig. 1, ANCOVA models revealed significant group differences across all linguistic and acoustic features, except for pitch variance. In terms of linguistic features, individuals with MDD used language that was more negatively valenced, emotionally intense, and lower in agency. In terms of acoustic features, individuals with MDD exhibited lower pitch, more time spent pausing, and slower speech rate. The speech composite score also differed significantly between groups. All results remained statistically significant after applying FDR

correction for multiple comparisons. Sensitivity analyses, which further adjusted for age at which English was learned and assessment location, yielded similar findings (see Table S1).

In a post hoc ROC analysis, we examined whether speech features could distinguish between individuals with MDD and non-depressed control participants. Like the analyses above, all speech features exhibited moderate to strong discriminability for depression, with the exception of pitch variance. The speech composite score demonstrated the greatest discriminability with an area under the curve of 0.90 (TableS2).

Question 2: Are there associations between speech features and depression severity, as measured by the HAMD-17?

As summarized in Table S3, the only linguistic feature to show a significant association with depression severity was agency, such that lower agency was associated with greater depression severity. However, this relationship did not survive FDR correction. None of the acoustic features were significantly associated with depression severity, nor was the composite speech score. Sensitivity analyses, which further adjusted for age at which English was learned and assessment location, yielded similar results (Table S4). Plots depicting these relationships are provided in Figure S2 and full model output including results from the sensitivity analysis is provided in Table S4.

Question 3. Are speech features associated with executive function, as measured by the IPAST?

Before examining associations between speech features and anti-saccade RT on the IPAST, we assessed whether anti-saccade RT differed between individuals with MDD and non-depressed control participants. Analyses adjusted for age, sex, years of education, and pro-saccade RT to account for individual differences in processing speed. Results showed that individuals with MDD exhibited slower anti-saccade RT than control participants ($F(1, 73) = 5.4$, $p = 0.02$, $\eta^2 = 0.07$). However, depression severity (HAMD-17 scores) was not significantly associated with anti-saccade RT ($\beta = -0.001$ $t = -0.001$, $p = 0.999$, $f^2 = -0.01$).

While anti-saccade RT was not associated with depression severity, it was significantly associated with several speech features (Table 4; Fig. 2). After adjusting for covariates, including pro-saccade RT, all linguistic features were significantly associated with slower anti-saccade RT, including language that was more negatively valenced, more emotionally intense, and less agentic. However, the association with emotionally intense language did not survive FDR correction. Among acoustic features, lower mean pitch was significantly associated with slower anti-saccade RT and survived FDR correction. More time spent pausing and slower speech rate did not reach statistical significance, though both showed small effect sizes. Pitch variance was not significantly associated with anti-saccade RT. The composite speech score was significantly and negatively associated with anti-saccade RT and survived FDR correction. Sensitivity analyses, which further adjusted for age at which English was learned and assessment location, yielded similar findings (Table S5).

To assess whether these associations were specific to anti-saccade RT, we performed additional linear regression analyses to test whether speech features were related to pro-saccade RT. As shown in Table S6, slower speech rate was the only significant association, but this effect did not survive FDR correction.

Finally, given the observed associations between speech features and anti-saccade RT, we examined whether the differences in speech features between individuals with MDD and controls could be explained by anti-saccade RT, which also differed between the groups. We used ANCOVA models to compare speech features between the two groups, adjusting for age, sex, years of education, audio recording duration, and anti-saccade RT. Group differences in language valence, emotional intensity, mean pitch, time spent pausing, speech rate, and the

Table 2. Demographic and clinical characteristics of participants.

	Participants with MDD	Non-depressed control participants
N	66	54
Age, mean (SD)	45.2 (15.1)	44.5 (17.2)
Sex, no (% female)	36 (54.5)	41 (75.9)*
HAMD-17, mean (SD)[1]	21.4 (4.2)	NA
Education, mean years (SD)	15.2 (2.3)	17.0 (1.8)***
Age English learned, mean years (SD)	2.3 (5.5)	2.5 (4.8)
Ethnicity, no (%)		
White	36 (54.5)	20 (37.0)
South Asian	3 (4.5)	11 (20.4)
East Asian	1 (1.5)	10 (18.5)
Arab	4 (6.1)	1 (1.9)
Hispanic	3 (4.5)	0 (0)
Filipino	1 (1.5)	3 (5.6)
Black Caribbean/ Black African	0 (0)	2 (3.7)
Mixed Ethnicity	7 (10.6)	6 (11.1)
Missing Data	11 (16.7)	1 (1.9)

[1]Two participants were missing HAMD-17 (Hamilton Depression Rating Scale) data.
***$p \le 0.001$; *$p \le 0.05$.

Table 3. Model output comparing speech features in individuals with MDD and non-depressed control participants.

Speech Feature	MDD Mean (SD)	Control Mean (SD)	*F*	*p*	*FDR adjusted p*	η^2
Valence	5.7 (0.6)	6.2 (0.4)	20.7	<0.001***	<0.001***	0.05
Emotional Intensity	4.1 (0.3)	3.9 (0.2)	11.8	<0.001***	<0.001***	0.03
Agency	5.5 (0.3)	5.7 (0.2)	15.8	<0.001***	<0.001***	0.02
Mean Pitch (Hz)	141.0 (36.2)	167.6 (33.3)	10.3	0.002**	0.002**	0.05
Pitch Variance (Hz)	3679 (4286)	3370 (2364)	0.3	0.32	0.57	0.003
Pauses	0.44 (0.2)	0.25 (0.2)	11.4	0.001***	0.001***	0.08
Speech Rate (words/min)	96.0 (44.8)	134.9 (40.2)	21.6	<0.001***	<0.001***	0.15
Speech Composite	0.29 (0.5)	−0.35 (0.3)	42.5	<0.001***	<0.001***	0.27

MDD major depressive disorder, *SD* standard deviation, *FDR* false discovery rate, *Hz* hertz.
***$p \le 0.001$; **$p \le 0.01$; *$p \le 0.05$.

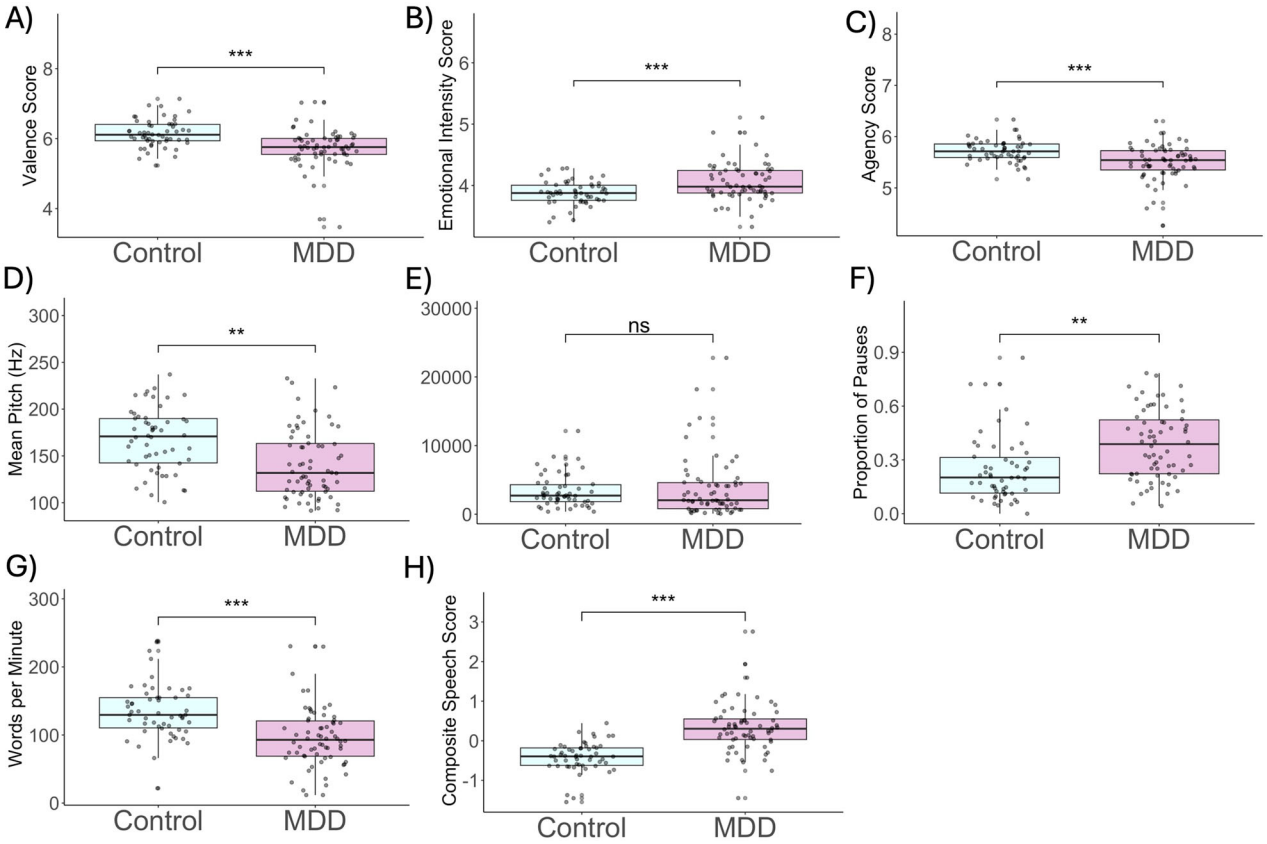

Fig. 1 Differences in speech features between individuals with Major Depressive Disorder (MDD) and non-depressed controls. Relative to controls, individuals with depression (**A**) used significantly more negatively valenced language ($F = 20.7$, FDR-adjusted $p < 0.001$, $n^2 = 0.05$), (**B**) used language that was significantly more emotionally intense ($F = 11.8$, FDR-adjusted $p < 0.001$, $n^2 = 0.05$), (**C**) exhibited significantly lower agency in their language ($F = 15.8$, $p < 0.001$, $n^2 = 0.02$), (**D**) spoke with a significantly lower mean pitch ($F = 10.3$, FDR-adjusted $p = 0.002$, $n^2 = 0.03$), (**E**) did not differ in pitch variance ($F = 0.3$, FDR-adjusted $p = 0.57$, $n^2 = 0.003$). **F** paused significantly more ($F = 11.4$, $p = 0.001$, $n^2 = 0.08$), (**G**) spoke significantly slower ($F = 21.6$, FDR-adjusted $p < 0.001$, $n^2 = 0.15$), and (**H**) had significantly higher speech composite scores ($F = 42.5$, FDR-adjusted $p < 0.001$, $n^2 = 0.27$). Abbreviations. MDD major depressive disorder. ***$p \le 0.001$; **$p \le 0.01$; *$p \le 0.05$.

Table 4. Associations between speech features and anti-saccade RT.

Speech Feature	Estimate *(β)*	*SE*	*t*	*p*	*FDR adjusted p*	f^2	Adjusted R^2
Valence	−17.31	6.36	−2.71	0.008**	0.02*	0.10	0.54
Emotional Intensity	25.30	11.89	2.13	0.04*	0.06	0.06	0.53
Agency	−33.06	10.95	−3.02	0.004**	0.01*	0.13	0.55
Mean Pitch (Hz)	−0.40	0.15	−2.70	0.009**	0.02*	0.10	0.54
Pitch Variance (Hz)	−0.001	0.001	−0.94	0.35	0.35	0.01	0.50
Pauses	36.65	19.67	1.86	0.07	0.09	0.05	0.52
Speech Rate (words/min)	−0.13	0.08	−1.67	0.10	0.12	0.04	0.51
Speech Composite	22.26	6.00	3.54	0.001***	0.006**	0.18	0.57

RT reaction time, *SE* standard error, *FDR* false discovery rate, *Hz* hertz.
***, $p \le 0.001$; **$p \le 0.01$; *$p \le 0.05$.

composite speech score remained significant, whereas agency was no longer significant (Table S7). These findings suggest that speech differences between groups are not fully attributable to differences in executive function, as measured by anti-saccade RT.

DISCUSSION

The present study examined whether speech could serve as an objective indicator of depressive symptoms and associated cognitive difficulties in individuals with MDD. There were three main findings. First, consistent with prior research, we found that both linguistic and acoustic speech features [7, 8], as well as performance on an eye-tracking test of executive function [52], differed between individuals with MDD and non-depressed control participants. Second, and contrary to our hypothesis, none of the speech features were associated with depression severity, as measured by the HAMD-17. Third, several speech features were associated with executive function performance, a domain commonly affected in depression [46].

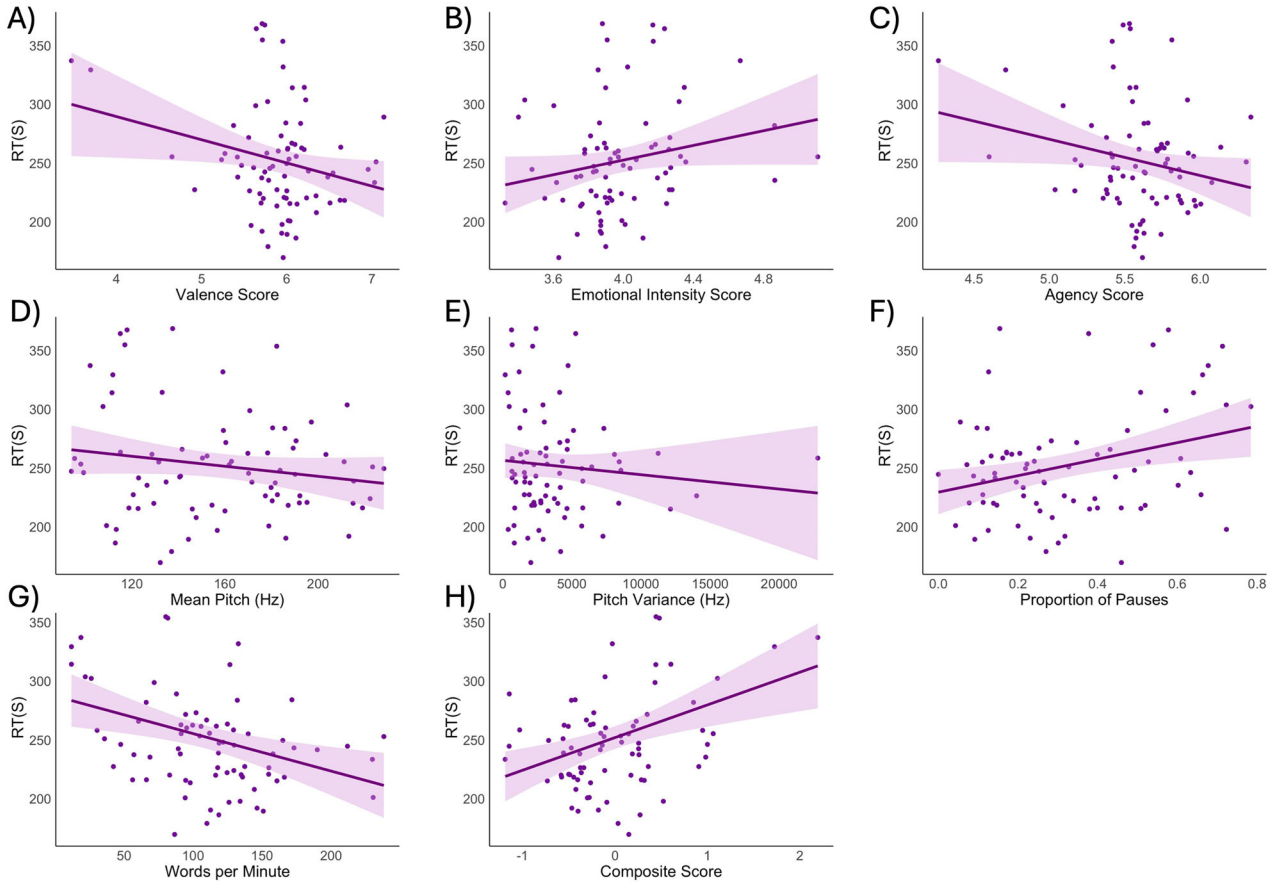

Fig. 2 Associations between speech features and performance on the executive function task, as measured by anti-saccade reaction time. A More negatively valenced language was associated with worse performance (FDR-adjusted $p = 0.02$, $f^2 = 0.10$). **B** More emotionally intense language was associated with worse performance at trend-level (FDR-adjusted $p = 0.06$, $f^2 = 0.06$). **C** Language reflecting less agency was associated with worse performance (FDR-adjusted $p = 0.01$, $f^2 = 0.13$). **D** Lower mean pitch was related to worse performance (FDR-adjusted $p = 0.02$, $f^2 = 0.10$). **E** There was no association between pitch variance and performance (FDR-adjusted $p = 0.35$, $f^2 = 0.10$). **F** More time spent pausing was associated with worse performance (FDR-adjusted $p = 0.09$, $f^2 = 0.05$). **G** Slower speech rate was associated with worse performance at trend level (FDR-adjusted $p = 0.12$, $f^2 = 0.04$). **H** Greater speech composite scores were associated with worse performance (FDR-adjusted $p = 0.006$, $f^2 = 0.18$). RT reaction time (slower reaction time reflects worse executive function performance), s Seconds.

Our findings build on prior research suggesting that linguistic and acoustic speech features differ between individuals with depression and non-depressed individuals [5, 6, 9]. Specifically, in the present study, we found that individuals with MDD used language that was more negatively valenced, higher in emotional intensity, and lower in agency. They also exhibited lower pitch, a slower speech rate, and more time spent pausing. Notably, among the individual speech features, slower speech rate exhibited the largest effect size ($\eta^2 = 0.15$), whereas the composite speech score demonstrated the strongest overall effect ($\eta^2 = 0.27$).

Consistent with these group differences, ROC analyses showed that the same speech features provided moderate to strong discrimination of MDD, with the speech composite achieving the highest classification accuracy (AUC = 0.90). Pitch variance was the only feature that did not differ between groups and showed weak classification accuracy. Together, these findings highlight the potential of speech-based measures as objective and accessible indicators of depressive symptoms.

Although speech features differed between individuals with MDD and non-depressed control participants, after FDR correction none were significantly associated with depression severity, as measured by the HAMD-17. Several factors may account for this lack of association. One possibility is that speech features serve more effectively as markers of depression presence than severity, given that they differed between individuals with MDD and controls and discriminated between the two groups. This interpretation is consistent with prior research suggesting that cognitive difficulties are commonly observed in individuals with depression but often fail to correlate with severity scores on traditional depression scales [67].

Another possibility is that speech features capture aspects of depression that the HAMD-17 does not. The HAMD-17 predominantly assesses somatic symptoms such as fatigue, insomnia, sexual dysfunction, and appetite/weight changes [68], with less attention placed on cognitive and behavioral components of depression, such as indecisiveness and psychomotor slowing [69–71]. It is possible that depression scales that more fully capture these symptoms may reveal stronger associations between speech features and depression severity.

Although speech features were not associated with depression severity, they were associated with executive function performance, as measured by anti-saccade RT on the IPAST. Slower anti-saccade RT was associated with more negative and less agentic language, lower mean pitch, and the composite speech score. Associations with more emotionally intense language and more time spent pausing showed similar patterns but did not meet the threshold for statistical significance. These findings suggest that speech features may capture executive difficulties not fully

reflected by the HAMD-17, highlighting a potential limitation of the scale. Importantly, these associations persisted after controlling for processing speed, and speech features were not related to processing speed, indicating that the relationships were driven by executive difficulties rather than general slowing.

The observed associations between linguistic speech features (i.e., more negatively valenced and less agentic language) and poorer executive function align with cognitive models of depression. These models propose that executive dysfunction contributes to difficulties suppressing negative information [48, 49, 72], reduces motivation and goal-directed behavior [72], and promotes negative self-appraisals [47, 73]. We also found that lower mean pitch was associated with poorer executive function performance. This may reflect diminished top-down executive control of expressive behavior, resulting in flatter and less varied speech.

Although slower speech and more time spent pausing have been attributed to psychomotor slowing in depression [67], we did not observe significant associations between these features and pro-saccade or anti-saccade RT after correcting for multiple comparisons. One possibility is that pro-saccade latency primarily captures basic sensorimotor speed, whereas speech rate and pausing rely to a greater extent on higher-order processes such as working memory, attentional control, and language planning [74–76]. Speech rate and pausing were also unrelated to anti-saccade RT, and may capture processes that are distinct from the executive demands required for fluent and coherent speech. Together, these results suggest that aspects of speech may serve as sensitive markers of executive dysfunction in depression, reflecting cognitive difficulties not captured by standard clinical assessments.

The study findings should be interpreted within the context of its limitations. First, our sample was limited to individuals with treatment-resistant MDD, all with HAMD-17 scores above 16 (moderate severity). This constrains the generalizability of our findings, leaving it unclear whether they apply to individuals with non–treatment-resistant MDD or to those with milder depressive symptoms. Second, HAMD-17 data were not collected for the non-depressed control group, precluding examination of associations across the full sample. Including this measure in both groups might have strengthened the ability to detect relationships between HAMD-17 scores and speech features. Third, the eye-tracking task was completed by a smaller subset of participants due to scheduling constraints and the need for in-person assessment. Importantly, those who completed the task did not differ significantly from those who did not in terms of age, sex, years of education, age of English acquisition, and HAMD-17 scores (Table S8). Fourth, the cross-sectional study design precluded us from determining whether speech features can track changes in depression severity over time, including response to treatment. Fifth, our selection of speech features was hypothesis-driven, focusing on a predefined subset of linguistic and acoustic features previously associated with depression. While this allowed us to test theory-driven predictions, it may have led to the omission of other potentially informative features. Future studies with larger samples could benefit from a more exploratory or data-driven approach to identify additional relevant markers. Finally, we used the HAMD-17 to assess depression severity, as it is widely used, but we have also noted its limitations. This means we are validating our findings against a measure we believe may be incomplete, highlighting a broader challenge in the field.

In conclusion, speech features differed significantly between individuals with MDD and non-depressed control participants and were associated with executive function performance. These findings suggest that speech features could serve as an objective and valuable tool for assessing both depression and its associated cognitive difficulties, an important consideration given the strong link between cognitive difficulties with poor functional outcomes and reduced quality of life [4]. Moreover, speech markers are low-burden, cost-effective, and can be collected remotely, making them a highly practical tool for both clinical and research settings.

DATA AVAILABILITY

The datasets generated and analyzed in this study are available upon reasonable request from the corresponding author.

REFERENCES

1. Depressive disorder (depression). https://www.who.int/news-room/fact-sheets/detail/depression. Updated August, 2025. Accessed 2 May 2025.
2. Lou Y, Lei Y, Mei Y, Leppänen PHT, Li H Review of abnormal self-knowledge in major depressive disorder. Front Psychiatry 2019;10. https://doi.org/10.3389/fpsyt.2019.00130.
3. Blaney PH. Affect and memory: a review. Psychol Bull. 1986;99:229–46.
4. McIntyre RS, Cha DS, Soczynska JK, Woldeyohannes HO, Gallaugher LA, Kudlow P, et al. Cognitive deficits and functional outcomes in major depressive disorder: determinants, substrates, and treatment interventions. Depress Anxiety. 2013;30:515–27.
5. Cummins N, Sethu V, Eppsgspan>J, Schnieder S, Krajewski J. Analysis of acoustic space variability in speech affected by depression. Speech Commun. 2015;75:27–49.
6. Low DM, Bentley KH, Ghosh SS. Automated assessment of psychiatric disorders using speech: a systematic review. Laryngoscope Investig Otolaryngol. 2020;5:96–116.
7. Cowie R, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias S, Fellenz W, et al. Emotion recognition in human-computer interaction. IEEE Signal Process Mag. 2001;18:32–80.
8. Yang B, Lugger M. Emotion recognition from speech signals using new harmony features. Signal Process. 2010;90:1415–23.
9. Tølbøll KB. Linguistic features in depression: a meta-analysis. J Lang Work - Sprogvidenskabeligt Stud. 2019;4:39–59.
10. Burkhardt HA, Alexopoulos GS, Pullmann MD, Hull TD, Areán PA, Cohen T. Behavioral activation and depression symptomatology: longitudinal assessment of linguistic indicators in text-based therapy sessions. J Med Internet Res. 2021;23:e28244.
11. Rude S, Gortner E-M, Pennebaker J. Language use of depressed and depression-vulnerable college students. Cogn Emot. 2004;18:1121–33.
12. Lyons M, Aksayli ND, Brewer G. Mental distress and language use: linguistic analysis of discussion forum posts. Comput Hum Behav. 2018;87:207–11.
13. Trifu RN, Nemeș B, Herta DC, Bodea-Hategan C, Talaș DA, Coman H Linguistic markers for major depressive disorder: a cross-sectional study using an automated procedure. Front Psychol 2024;15. https://doi.org/10.3389/fpsyg.2024.1355734.
14. Stade EC, Ungar L, Eichstaedt JC, Sherman G, Ruscio AM. Depression and anxiety have distinct and overlapping language patterns: results from a clinical interview. J Psychopathol Clin Sci. 2023;132:972–83.
15. Baddeley JL, Pennebaker JW, Beevers CG. Everyday social behavior during a major depressive episode. Soc Psychol Personal Sci. 2013;4:445–52.
16. De Choudhury M, Counts S, Horvitz E Social media as a measurement tool of depression in populations. In: *Proceedings of the 5th Annual ACM Web Science Conference*. New York, NY, USA: Association for Computing Machinery; 2013. pp 47–56.
17. Zogan H, Razzak I, Wang X, Jameel S, Xu G. Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media. World Wide Web. 2022;25:281–304.
18. Lihao Wang. Social media data mining and mental health status assessment based on sentiment analysis. J Electr Syst. 2024;20:381–6.
19. Settanni M, Marengo D Sharing feelings online: studying emotional well-being via automated text analysis of facebook posts. Front Psychol 2015;6. https://doi.org/10.3389/fpsyg.2015.01045.
20. Howes C, Purver M, McCabe R Linguistic indicators of severity and progress in online text-based therapy for depression. In: Resnik P, Resnik R, Mitchell M (eds). *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. Baltimore, Maryland, USA: Association for Computational Linguistics: 2014. pp 7–16.
21. Breznitz Z. Verbal indicators of depression. J Gen Psychol. 1992;119:351–63.
22. Darby JK, Simmons N, Berger PA. Speech and voice parameters of depression: a pilot study. J Commun Disord. 1984;17:75–85.
23. Kuny S, Stassen HH. Speaking behavior and voice sound characteristics in depressive patients during recovery. J Psychiatr Res. 1993;27:289–307.

24. Nilsonne Å. Acoustic analysis of speech variables during depression and after improvement. Acta Psychiatr Scand. 1987;76:235–45.
25. Kiss G, Vicsi K. Mono- and multi-lingual depression prediction based on speech processing. Int J Speech Technol. 2017;20:919–35.
26. Alghowinem S. From joyous to clinically depressed: mood detection using multimodal analysis of a person's appearance and speech. *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*; 2013; Geneva, Switzerland. p. 648–654. https://doi.org/10.1109/ACII.2013.113.
27. Calić G, Petrović-Lazić M, Mentus T, Babac S. Acoustic features of voice in adults suffering from depression. Psihol Istraživanja. 2022;25:183–203.
28. 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. Proc Interspeech 2011; 3001-4. https://doi.org/10.21437/Interspeech.2011-751
29. Darby JK, Hollien H. Vocal and speech patterns of depressive patients. Folia Phoniatr. 1977;29:279–91.
30. Godfrey HP, Knight RG. The validity of actometer and speech activity measures in the assessment of depressed patients. Br J Psychiatry. 1984;145:159–63.
31. Greden JF, Carroll BJ. Decrease in speech pause times with treatment of endogenous depression. Biol Psychiatry. 1980;15:575–87.
32. Hardy P, Jouvent R, Widlöcher D. Speech pause time and the retardation rating scale for depression (ERD): towards a reciprocal validation. J Affect Disord. 1984;6:123–7.
33. Szabadi E, Bradshaw CM, Besson JA. Elongation of pause-time in speech: a simple, objective measure of motor retardation in depression. Br J Psychiatry J Ment Sci. 1976;129:592–7.
34. 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.
35. Menne F, Dörr F, Schräder J, Tröger J, Habel U, König A, et al. The voice of depression: speech features as biomarkers for major depressive disorder. BMC Psychiatry. 2024;24:794.
36. Alpert M, Pouget ER, Silva RR. Reflections of depression in acoustic measures of the patient's speech. J Affect Disord. 2001;66:59–69.
37. 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 International Conference on Body Sensor Networks*; 2013; Cambridge, MA, USA. p. 1–6. https://doi.org/10.1109/BSN.2013.6575522.
38. Quatieri TF, Malyska N. Vocal-source biomarkers for depression: a link to psychomotor activity. *Proceedings of the 13th Annual Conference of the International Speech Communication Association (Interspeech 2012)*; 2012 Sep 9–13; Portland, OR, USA. p. 1059–1062. https://doi.org/10.21437/Interspeech.2012-311.
39. Mundt JC, Vogel AP, Feltner DE, Lenderking WR. Vocal acoustic biomarkers of depression severity and treatment response. Biol Psychiatry. 2012;72:580–7.
40. Ellgring H, Scherer KR. Vocal indicators of mood change in depression. J Nonverbal Behav. 1996;20:83–110.
41. Cannizzaro M, Harel B, Reilly N, Chappell P, Snyder PJ. Voice acoustical measurement of the severity of major depression. Brain Cogn. 2004;56:30–35.
42. Stassen HH, Kuny S, Hell D. The speech analysis approach to determining onset of improvement under antidepressants. Eur Neuropsychopharmacol J Eur Coll Neuropsychopharmacol. 1998;8:303–10.
43. 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 Neurolinguist. 2007;20:50–64.
44. Alhanai T, Au R, Glass J. Spoken language biomarkers for detecting cognitive impairment. *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*; 2017 Dec 16–20; Okinawa, Japan. p. 409–416. https://doi.org/10.1109/ASRU.2017.8268965.
45. Toth L, Hoffmann I, Gosztolya G, Vincze V, Szatloczki G, Banreti Z, et al. A Speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. Curr Alzheimer Res. 2018;15:130–8.
46. Lee RSC, Hermens DF, Porter MA, Redoblado-Hodge MA. A meta-analysis of cognitive deficits in first-episode major depressive disorder. J Affect Disord. 2012;140:113–24.
47. Gotlib IH, Joormann J. Cognition and depression: current status and future directions. Annu Rev Clin Psychol. 2010;6:285–312.
48. LeMoult J, Gotlib IH. Depression: a cognitive perspective. Clin Psychol Rev. 2019;69:51–66.
49. Grahek I, Everaert J, Krebs RM, Koster EHW. Cognitive control in depression: toward clinical models informed by cognitive neuroscience. Clin Psychol Sci. 2018;6:464–80.
50. Rock PL, Roiser JP, Riedel WJ, Blackwell AD. Cognitive impairment in depression: a systematic review and meta-analysis. Psychol Med. 2014;44:2029–40.
51. Buist-Bouwman MA, Ormel J, de Graaf R, de Jonge P, van Sonderen E, Alonso J, et al. Mediators of the association between depression and role functioning. Acta Psychiatr Scand. 2008;118:451–8.
52. Yep R, Smorenburg ML, Riek HC, Calancie OG, Kirkpatrick RH, Perkins JE et al. Interleaved pro/anti-saccade behavior across the lifespan. Front Aging Neurosci 2022;14. https://doi.org/10.3389/fnagi.2022.842549.
53. Berlim MT, Turecki G. Definition, assessment, and staging of treatment-resistant refractory major depression: a review of current concepts and methods. Can J Psychiatry Rev Can Psychiatr. 2007;52:46–54.
54. Winterlight Labs. https://winterlightlabs.com/. Updated 2025, Accessed 6 Feb 2023.
55. Robin J, Xu M, Balagopalan A, Novikova J, Kahn L, Oday A, et al. Automated detection of progressive speech changes in early alzheimer's disease. Alzheimers Dement Amst. 2023;15:e12445.
56. Heiberger VL, Horii Y Jitter and shimmer in sustained phonation. In: Lass NJ (ed). *Speech and Language*. Elsevier, 1982, pp 299–332.
57. Farrús M, Hernando J, Ejarque P Jitter and shimmer measurements for speaker recognition. In: *Interspeech 2007*. ISCA, 2007, pp 778-81.
58. Mitra V, Franco H, Stern RM, van Hout J, Ferrer L, Graciarena M et al. Robust features in deep-learning-based speech recognition. In: Watanabe S, Delcroix M, Metze F, Hershey JR, editors., New Era for Robust Speech Recognition: Exploiting Deep Learning. Cham: Springer International Publishing; 2017. pp. 187–217. pp.
59. Al-Karawi KA, Al-Bayati B. The effects of distance and reverberation time on speaker recognition performance. Int J Inf Technol. 2024;16:3065–71.
60. Coe BC, Munoz DP. Mechanisms of saccade suppression revealed in the anti-saccade task. Philos Trans R Soc Lond B Biol Sci. 2017;372:20160192.
61. McDowell JE, Dyckman KA, Austin BP, Clementz BA. Neurophysiology and neuroanatomy of reflexive and volitional saccades: evidence from studies of humans. Brain Cogn. 2008;68:255–70.
62. Kahana Levy N, Lavidor M, Vakil E. Prosaccade and antisaccade paradigms in persons with alzheimer's disease: a meta-analytic review. Neuropsychol Rev. 2018;28:16–31.
63. Cecchetti S, Duchowski AT, Cavallo M. Eye-tracking metrics as a digital biomarker for neurocognitive disorders in multiple sclerosis: a scoping review. Brain Sci. 2025;15:149.
64. Riek HC, Brien DC, Coe BC, Huang J, Perkins JE, Yep R, et al. Cognitive correlates of antisaccade behaviour across multiple neurodegenerative diseases. Brain Commun. 2023;5:fcad049.
65. Poletti B, Carelli L, Solca F, Lafronza A, Pedroli E, Faini A, et al. An eye-tracker controlled cognitive battery: overcoming verbal-motor limitations in ALS. J Neurol. 2017;264:1136–45.
66. Coe BC, Huang J, Brien DC, White BJ, Yep R, Munoz DP. Automated analysis pipeline for extracting saccade, pupil, and blink parameters using video-based eye tracking. Vision. 2024;8:14.
67. McClintock SM, Husain MM, Greer TL, Cullum CM. Association between depression severity and neurocognitive function in major depressive disorder: a review and synthesis. Neuropsychology. 2010;24:9–34.
68. Fried EI. The 52 symptoms of major depression: lack of content overlap among seven common depression scales. J Affect Disord. 2017;208:191–7.
69. Maust D, Cristancho M, Gray L, Rushing S, Tjoa C, Thase ME. Psychiatric rating scales. Handb Clin Neurol. 2012;106:227–37.
70. Rabin JS, Davidson B, Giacobbe P, Hamani C, Cohn M, Illes J, et al. Neuromodulation for major depressive disorder: innovative measures to capture efficacy and outcomes. Lancet Psychiatry. 2020;7:1075–80.
71. Rosenblat JD, Simon GE, Sachs GS, Deetz I, Doederlein A, DePeralta D, et al. Treatment effectiveness and tolerability outcomes that are most important to individuals with bipolar and unipolar depression. J Affect Disord. 2019;243:116–20.
72. Roiser JP, Sahakian BJ. Hot and cold cognition in depression. CNS Spectr. 2013;18:139–49.
73. Grahek I, Shenhav A, Musslick S, Krebs RM, Koster EHW. Motivation and cognitive control in depression. Neurosci Biobehav Rev. 2019;102:371–81.
74. Sjerps MJ, Meyer AS. Variation in dual-task performance reveals late initiation of speech planning in turn-taking. Cognition. 2015;136:304–24.
75. Eichorn N, Pirutinsky S, Marton K. Effects of different attention tasks on concurrent speech in adults who stutter and fluent controls. J Fluen Disord. 2019;61:105714.
76. Shen C, Janse E. Maximum speech performance and executive control in young adult speakers. J Speech Lang Hear Res JSLHR. 2020;63:3611–27.
77. Warriner AB, Kuperman V, Brysbaert M. Norms of valence, arousal, and dominance for 13,915 english lemmas. Behav Res Methods. 2013;45:1191–207.

ACKNOWLEDGEMENTS

We thank all study participants and the authors for their invaluable contributions to study design, data collection, and manuscript preparation. We also appreciate Winterlight Labs for providing access to their speech-analysis platform. This work was

funded by the Harquail Centre for Neuromodulation and the Canadian Institutes of Health Research (CIHR, 191686).

AUTHOR CONTRIBUTIONS

MW and JSR conceived and designed the study. MW, MAW, GG, and LP collected the speech data; RY and CP collected the eye-tracking data. MW performed data analysis and interpretation with input from JSR and MV. MW wrote the manuscript with critical revisions from JSR. DPM, BCC, and DB developed the eye-tracking task and assisted with preprocessing. JR, MJS, and WS provided guidance on the use of Winterlight Labs' speech analysis tool. SN and PG screened participants, conducted clinical assessments, and contributed to study design and interpretation. NL, as head of the Harquail Centre, facilitated patient recruitment and provided input on study design and interpretation. YY served as consulting speech pathologist. All authors reviewed and approved the final manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41398-025-03728-2.

Correspondence and requests for materials should be addressed to Jennifer S. Rabin.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.