



Published in final edited form as:

J Affect Disord. 2023 February 15; 323: 675–678. doi:10.1016/j.jad.2022.12.047.

Word usage in spontaneous speech as a predictor of depressive symptoms among youth at high risk for mood disorders

Marc J. Weintraub^{a,*}, Filippo Posta^b, Megan C. Ichinose^a, Armen C. Arevian^c, David J. Miklowitz^a

^aUCLA Semel Institute for Neuroscience and Human Behavior, Los Angeles, CA, United States of America

^bEstrella Mountain Community College, Avondale, AZ, United States of America

^cChorus Innovations, Long Beach, CA, United States of America

Abstract

Background: We examined whether digital phenotyping of spontaneous speech, such as the use of specific word categories during speech samples, was associated with depressive symptoms in youth who were at familial and clinical risk for mood disorders.

Methods: Participants (ages 13–19) had active mood symptoms, mood instability, and at least one parent with bipolar or major depressive disorder. During a randomized trial of family-focused therapy, participants were instructed to make weekly calls to a central voice server and leave speech samples in response to automated prompts. We coded youths' speech samples with the Linguistic Inquiry and Word Count system and used machine learning to identify the combination of speech features that were most closely associated with the course of depressive symptoms over 18 weeks.

Results: A total of 253 speech samples were collected from 44 adolescents (mean age = 15.8 years; SD = 1.6) over 18 weeks. Speech containing affective processes, social processes, drives toward risk or reward, nonfluencies, and time orientation words were correlated with depressive symptoms at concurrent time periods ($p < 0.01$). Machine learning analyses revealed that affective processes, nonfluencies, drives and risk words combined to most strongly predict changes in depressive symptoms over 18 weeks of treatment.

Limitations: Study results were limited by the small sample and the exclusion of paralinguistic or contextual variables in analyzing speech samples.

*Corresponding author at: UCLA Semel Institute for Neuroscience and Human Behavior, Department of Psychiatry, 760 Westwood Plaza, A8-259, Los Angeles, CA 90095, United States of America. mjweintraub@mednet.ucla.edu (M.J. Weintraub).

Contributors

Drs. Miklowitz (Principal Investigator) and Dr. Posta had full access to all study data and take responsibility for the integrity of the data and accuracy of the data analyses. Dr. Posta was the lead statistician. Dr. Arevian created the Chorus web platform on which the mobile app was developed. Drs. Miklowitz, Arevian and Walshaw contributed to the design of the study. Dr. Weintraub contributed to data management and analysis. Drs. Walshaw, Dr. Denenny, Dr. Weintraub, and Ms. Frey and Ms. Morgan-Fleming contributed to the acquisition and interpretation of study data. Ms. Wilkerson was the lead programmer on the mobile app and contributed to data management. Drs. Miklowitz and Arevian provided administrative and material support. All authors contributed to the drafting and critical revision of the manuscript, and have approved the final article.

Conclusions: In youth at high risk for mood disorders, knowledge of speech patterns may inform prognoses during outpatient psychosocial treatment.

Keywords

Machine learning; Linguistic; LIWC; Adolescents; Depression; Bipolar; Family-focused therapy

1. Introduction

A substantial body of work dating back to Kraepelin (1921) has found that patients with mood disorders have specific speech abnormalities (Cummins et al., 2015). Depressed patients tend to speak “slowly, hesitatingly, monotonously, sometimes stuttering...” (Kraepelin, 1921). Positive and negative emotion words and first-person singular words are associated with greater depressive symptoms in college students (Rude et al., 2004). The speech of adults and adolescents with depression has a greater frequency of references to the past compared to speech of healthy volunteers (Habermas et al., 2008; Jones et al., 2020).

Analysis of speech samples can elucidate cognitive and affective processes that underlie mood disorders and may be useful in determining prognosis or predicting treatment response. For example, speech samples may be useful in assessing suicide risk in patients (Cummins et al., 2015) or identifying mental health concerns in the general population using social media data (Uban et al., 2021). These strategies may help direct people to treatment and/or determine the intensity of recommended interventions.

Recent work has found that speech samples are feasible to collect longitudinally and can help track within-individual changes in mental status (Arevian et al., 2020). However, there is a paucity of longitudinal data on whether speech features relate to psychiatric health in youth at earlier stages of symptom development. In this study, we explored whether speech features were associated with depression ratings over 6 months among “high-risk” adolescents who had mood symptoms and at least one parent with an established mood disorder. Because participants were enrolled in a randomized clinical trial of psychological treatments, it was possible to obtain speech samples on a weekly basis using a centralized voice server. On an exploratory basis, we conducted a machine learning analysis to examine which speech features combined to best predict depression throughout the 18-week study as well as which features best predicted changes in depression over the 27-week study period.

2. Methods

2.1. Study participants

We recruited youth who met the following criteria: (1) ages 13 years to 19 years; (2) current mood symptoms, as indicated by a score ≥ 12 on the Young Mania Rating Scale (Young et al., 1978) and/or ≥ 9 on the Children’s Depression Rating Scale, Revised (Poznanski and Mokros, 1996); (3) evidence of mood instability, as indicated by either a score ≥ 6 on the 10-item Parent General Behavior Inventory for Mania (Youngstrom et al., 2008) or ≥ 20 on the parent-rated 20-item Children’s Affective Lability Scale (Gerson et al., 1996); and (4) at least one biological parent with a history of major depressive disorder or bipolar I or II

disorder, as indicated by the MINI International Neuropsychiatric Interview (Sheehan et al., 2010).

The youth and key relatives (usually parents) participated in a randomized trial family-focused therapy (12 sessions in 18 weeks) enhanced by one of two mobile apps (Miklowitz et al., 2021): FFT with MyCoachConnect (FFT-MCC), an educational and skill-oriented app, or FFT-Track, a mood tracking app. Both apps included reminders and a direct link to a “voice journal” for participants to leave a speech sample. The voice journal asked participants to respond to two prompts: (1) “What has been going on with your family over the past day or two. What went well? What stressed you out?” and (2) “Tell us something that went well or stressed you out about school, friends, or anything else in the last couple of days.” Participants were asked to leave at least one voice journal weekly without any instructions or limitations regarding the duration of time for which they should speak.

2.2. Clinical outcome assessments

Study assessors interviewed the adolescent and one parent at four time-points: baseline, 9 weeks (mid-treatment), 18 weeks (post-treatment), and 27 weeks (follow-up). Depressive symptom severity from the 9 previous weeks of the trial were rated at each study time-point using Psychiatric Status Ratings (PSRs) from the Adolescent Longitudinal Interval Follow-up Evaluation (Keller et al., 1987). Weekly PSRs for depression (the primary outcome for this study) were based on a consensus between the adolescent’s and parent’s reports and averaged over 9-week intervals, with averages ranging from 1 (no depressive symptoms) to 6 (severe symptoms). Interrater reliability (intraclass r) for PSR depression scores was 0.88.

2.3. Linguistic Inquiry and Word Count (LICW)

Each of the youth’s voice journal speech samples was transcribed and analyzed to derive speech feature data using the Linguistic Inquiry Word Count (LIWC) software (Pennebaker et al., 2015). The LIWC’s 93 speech features were pared down to 20 features that reflected affective processes, social processes, drives, informal, and time orientation words (see Table 1). The 20 speech features were selected by consensus agreement of the study researchers based on previous research regarding the relationship of various speech features to depressive symptoms and features that were thought to indicate common processes in depression (e.g., reward processing words). An example speech sample and LIWC analyses are presented in Table 2.

2.4. Data analysis

Speech samples with fewer than 25 words were removed. Speech data were then matched by date (time from random assignment to the date the participant produced the sample) with the corresponding 1–6 PSR depression score from the same week over 18 weeks of treatment. In cases where a participant had multiple speech samples within the same week, their speech data were aggregated into a single sample by taking the average of the 20 speech features. When PSR scores were missing for the week in which a speech sample was given, we imputed the average between the previous and following week.

Using Pearson correlations, we examined the associations between the 20 pre-selected speech features from each speech sample and the PSR depression score from the corresponding week. Next, we examined the relation of baseline speech data (i.e., the first voice journal produced after random assignment) to the change in PSR depression scores from baseline to 18 weeks. PSR depression change scores were calculated by subtracting the baseline depression score from the 18-week depression score.

We used machine learning to identify (1) the speech features that most strongly correlated with concurrent depressive symptoms over 18 weeks, and (2) the speech samples that most strongly predicted change in depressive symptoms from baseline to 18 weeks. We applied Support Vector Machine (SVM) algorithms to predict depression scores from the 20 pre-selected LIWC features. All results come from SVM algorithms modeled with radial (i.e., gaussian/bell-shaped) kernels, as those results were the most robust. The optimal speech feature subset was obtained through recursive feature elimination, which is an optimization algorithm that recursively removes the least predictive feature and randomly adds a previously removed feature to maximize the exploration of the predictor-space (Kuhn and Johnson, 2013).

3. Results

Of 53 adolescents enrolled in the trial, 44 adolescents had viable speech samples. The youth were 15.8 years of age ($SD = 1.6$), which included 30 females (68.2 %). The majority identified as white ($n = 32$; 72.7 %) and of non-Latinx ethnicity ($n = 34$; 77.3 %). The sample included 36 (81.8 %) with DSM-5 depressive spectrum disorders, including 27 with a single major depressive episode, 7 with recurrent major depressive episodes, and 2 with an other specified depressive disorder. The remaining 8 (18.2 %) youth had a DSM-5 bipolar spectrum disorder, including 3 with bipolar I disorder, 1 with bipolar II disorder and the 4 with unspecified bipolar disorder. Youth at baseline had moderately severe depressive symptoms as indicated by mean PSR depression scores ($M = 4.41$, $SD = 0.87$).

A total of 253 speech samples were collected from the 44 adolescents. Participants recorded an average of 5.8 speech samples ($SD = 4.7$) with speech samples averaging 264.2 words per sample ($SD = 198.6$) over the 18-month study. Participants were split evenly between the two treatment conditions (FFT-MCC: $n = 21$, FFT-Track: $n = 23$) with no differences in the number of speech samples or LIWC speech feature values between treatment groups. There were no relationships between word count and depressive symptom severity in speech samples with 25 words or more ($b = 0.00$, $SE = 0.003$, $p = 0.77$). Speech samples were provided less frequently as the study progressed ($b = -0.03$, $SE = 0.003$, $p < 0.001$) and this decrease was similar across treatment groups ($b = 0.003$, $SE = 0.004$, $p = 0.29$).

Of the 20 pre-selected LIWC speech features, 13 correlated with concurrent A-LIFE depression scores (i.e., depression scores rated at the same week the speech sample was recorded; see Table 1). Social process and present-focused speech features were the only two features that occurred more frequently in youth with higher depression scores. Past-focused and leisure speech features were most strongly negatively associated with depression scores. Both negative and positive emotion speech features were negatively

correlated with depressive symptoms over the study period, indicating that participants who expressed more emotion words were less depressed over the 18-week study period. Within the drives word category, reward and risk words were negatively associated with depressive symptoms. Additionally, nonfluencies were negatively correlated with depression scores. Words related to family or friends (sub-features of social processes) or future orientation were not associated with concurrent depression scores.

Machine learning analyses were conducted on the 253 speech samples to examine what combination of features was mostly strongly correlated with PSR depression scores over 18 weeks. After entering the hypothesized 20 LIWC features into the model, five features emerged as strongly correlated with depression scores: affective processes, drives, informal, leisure, and risk ($r = 0.47$, 95 % CI: 0.37–0.56, $R^2 = 0.12$).

Next, we examined each participant's first speech sample to determine which of 20 LIWC features were most strongly associated with depression change scores from baseline to 18 weeks. Four features combined to predict changes most strongly in depression: affective processes, nonfluencies, drives and risk ($r = 0.68$, 95 % CI: 0.48–0.81, $R^2 = 0.11$).

4. Discussion

This study examined speech samples collected from youth with depressive spectrum or bipolar spectrum disorders who participated in a randomized trial of family-focused therapy. All participants had at least one parent with a lifetime history of mood disorder. We analyzed the relationships between speech features and depression scores over the course of an 18-week treatment and conducted an exploratory machine learning analysis to elucidate speech features that predicted change in depressive symptoms. Speech that included more affective processes, social processes, drives, nonfluencies, and time orientation words was correlated with depressive symptom severity at concurrent time periods. Machine learning analyses revealed that affective words, nonfluencies, and drive and risk words were most strongly predictive of changes in depressive symptoms from the beginning to the end of the 18-week study.

Study results indicated that greater use of both positive and negative emotion words was associated with improvements in depressive symptoms. This finding is divergent from previous work that has found that individuals with depression use more negative words than non-depressed individuals (Rude et al., 2004). A limitation of the LIWC method is that it measures the frequency of particular words in word categories independent of context. Thus, it is possible for negative emotion words to be used in a positive context (e.g., "I'm not hurt by what he said."). In cases like this, the word, "hurt," would be categorized under negative emotions, but may not necessarily be indicative of a depressive state. Communication using negative and positive words may suggest greater emotional awareness – the ability to identify and communicate one's emotions (Kranzler et al., 2016). Emotional blunting (i.e., the inability to feel positive negative emotions), a common feature in depression (Christensen et al., 2022), may also help explain why increased use of emotional words was associated with lower severity of depressive symptoms.

Greater use of drive words was associated with lower depressive symptoms. Similarly, sub-features of the personal concern word category (leisure and home words) were associated with lower depressive symptoms. Adolescents who are depressed or have parents with depression have blunted neural (e.g., striatal) responses to reward, suggesting that reward processing is impaired across the depressive spectrum (Luking et al., 2016). Depression is also associated with risk-averse behaviors, such as avoidance and escape (Haskell et al., 2020). The present study adds the observation that greater use of words indicating drive, risk, or positive or negative affect may be associated with greater mood improvement among youth in a family treatment program.

Prior work has found that a greater focus on the past in life narratives distinguishes adults with clinical depression and adolescents with depressive symptoms from healthy volunteers (Habermas et al., 2008; Jones et al., 2020). Surprisingly, a greater use of past focus words was associated with less severe depressive symptoms in this study. However, the voice journal task required that adolescents relate events that occurred in the prior week. This response set may have increased the use of past-oriented words, especially in speech samples that described drive, risk-taking, or emotional reactions.

4.1. Limitations

Because the LIWC system only measures the presence of word features within speech samples, we were not able to assess these speech features nor the context in which words were spoken. The study's sample size and pool of speech samples were too small to draw firm conclusions from the machine learning analyses. We also did not have enough samples to ascertain how specific speech features (e.g., drive, affective processes) changed in relation to changes in specific depressive symptoms (e.g., mood, fatigue, loss of interests).

5. Conclusion

There was considerable overlap in speech features that correlated with depressive symptoms at concurrent time points and features that predicted depressive symptoms over 18-weeks. These results suggest that speech, as captured in phone calls to a voice server, may serve as behavioral indicators of depressive symptom states. With replicability and increased accuracy of machine learning results, digital phenotyping of speech features may become a useful tool for predicting concurrent and future depressive symptoms in high-risk youth.

Acknowledgments

We wish to thank the following individuals for providing administrative support and study treatments: Patricia Walshaw, PhD, Danielle Denenny, PhD, Alissa Ellis, PhD, Elizabeth Horstmann, MD, Sarah Marvin, PhD, Robert Suddath, MD, Cassidy Zanko, MD, Monica Done, MS, Gigi Laurin, Samantha Frey, Georga Morgan-Fleming.

Role of the funding source

This study was supported in part by grants R34 MH117200 (Drs. Miklowitz and Arevian) from the National Institute of Mental Health (NIMH), and funds from the Carl and Roberta Deutsch Foundation, Kayne Family Foundation, Danny Alberts Foundation, AIM for Youth Mental Health, and the Jewish Community Fund of Los Angeles. The funding sources had no role in the design or conduct of the study; collection, management, analysis and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Conflict of interest

Dr. Weintraub reports research support from NIMH K23MH124015, the American Psychological Foundation, and the Friends of the UCLA Semel Institute. Dr. Arevian is CEO of Chorus Innovations and is supported by the California State Center of Excellence for Behavioral Health SB 852. He has a financial interest in Insight Health Systems, Inc. and Arevian Technologies Inc. Dr. Miklowitz receives research support from the National Institute for Mental Health (NIMH), the Danny Alberts Foundation, the Attias Family Foundation, the Carl and Roberta Deutsch Foundation, the Kayne Family Foundation, AIM for Mental Health, Jewish Community Foundation of Los Angeles, and the Max Gray Fund; and book royalties from Guilford Press and John Wiley and Sons. The other authors have no declarations of interest.

References

- Arevian AC, Bone D, Malandrakis N, Martinez VR, Wells KB, Miklowitz DJ, Narayanan S, 2020. Clinical state tracking in serious mental illness through computational analysis of speech. *PLoS one* 15, e0225695. [PubMed: 31940347]
- Christensen MC, Ren H, Fagiolini A, 2022. Emotional blunting in patients with depression. Part I: clinical characteristics. *Ann. General Psychiatry* 21, 1–8.
- Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF, 2015. A review of depression and suicide risk assessment using speech analysis. *Speech Comm.* 71, 10–49.
- Gerson AC, Gerring JP, Freund L, Joshi PT, Capozzoli J, Brady K, Denckla MB, 1996. The Children's affective lability scale: a psychometric evaluation of reliability. *Psychiatry Res.* 65, 189–198. [PubMed: 9029668]
- Habermas T, Ott LM, Schubert M, Schneider B, Pate A, 2008. Stuck in the past: negative bias, explanatory style, temporal order, and evaluative perspectives in life narratives of clinically depressed individuals. *Depress. Anxiety* 25, E121–E132. [PubMed: 17957809]
- Haskell AM, Britton PC, Servatius RJ, 2020. Toward an assessment of escape/avoidance coping in depression. *Behav. Brain Res* 381, 112363. [PubMed: 31739002]
- Jones LS, Anderson E, Loades M, Barnes R, Crawley E, 2020. Can linguistic analysis be used to identify whether adolescents with a chronic illness are depressed? *Clin. Psychol. Psychother* 27, 179–192. [PubMed: 31840339]
- Keller MB, Lavori PW, Friedman B, Nielsen E, Endicott J, McDonald-Scott P, Andreasen NC, 1987. The longitudinal interval follow-up evaluation: a comprehensive method for assessing outcome in prospective longitudinal studies. *Arch. Gen. Psychiatry* 44, 540–548. [PubMed: 3579500]
- Kraepelin E, 1921. *Manic-depressive Insanity and Paranoia*. E. & S. Livingstone.
- Kranzler A, Young JF, Hankin BL, Abela JR, Elias MJ, Selby EA, 2016. Emotional awareness: a transdiagnostic predictor of depression and anxiety for children and adolescents. *J. Clin. Child Adolesc. Psychol* 45, 262–269. [PubMed: 25658297]
- Kuhn M, Johnson K, 2013. *Applied Predictive Modeling*. Springer.
- Luking KR, Pagliaccio D, Luby JL, Barch DM, 2016. Reward processing and risk for depression across development. *Trends Cogn. Sci* 20, 456–468. [PubMed: 27131776]
- Miklowitz DJ, Weintraub MJ, Posta F, Walshaw PD, Frey SJ, Morgan-Fleming GM, Wilkerson CA, Denenny DM, Arevian AA, 2021. Development and open trial of a technology-enhanced family intervention for adolescents at risk for mood disorders. *J. Affect. Disord* 281, 438–446. [PubMed: 33360365]
- Pennebaker JW, Boyd RL, Jordan K, Blackburn K, 2015. *The Development and Psychometric Properties of LIWC2015*.
- Poznanski EO, Mokros HB, 1996. *Children's Depression Rating Scale, Revised (CDRSR)*. Western Psychological Services Los Angeles.
- Rude S, Gortner E-M, Pennebaker J, 2004. Language use of depressed and depression-vulnerable college students. *Cognit. Emot* 18, 1121–1133.
- Sheehan DV, Sheehan KH, Shytle RD, Janavs J, Bannon Y, Rogers JE, Milo KM, Stock SL, Wilkinson B, 2010. Reliability and validity of the mini international neuropsychiatric interview for children and adolescents (MINI-KID). *J. Clin. Psychiatry* 71.

- Uban A-S, Chulvi B, Rosso P, 2021. An emotion and cognitive based analysis of mental health disorders from social media data. *Futur. Gener. Comput. Syst* 124, 480–494.
- Young R, Biggs J, Ziegler V, Meyer D, 1978. A rating scale for mania: reliability, validity and sensitivity. *Br. J. Psychiatry* 133, 429–435. [PubMed: 728692]
- Youngstrom EA, Frazier TW, Demeter C, Calabrese JR, Findling RL, 2008. Developing a 10-item mania scale from the parent general behavior inventory for children and adolescents. *J. Clin. Psychiatry* 69, 4319.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

LIWC speech features and their correlation with concurrent A-LIFE depression scores.

Speech feature	Abbreviation	Example words	A-LIFE depression correlation
Affective processes	affect	Happy, cried	-0.17
Positive emotion	posemo	Love, nice, sweet	-0.13
Negative emotion	negemo	Hurt, ugly, nasty	-0.10
Social processes	social	Mate, talk, they	0.12
Family	family	Daughter, dad, aunt	-0.07
Friends	friend	Buddy, neighbor	0.01
Drives	drives		-0.11
Affiliation	affiliation	Ally, friend, social	-0.06
Achievement	achieve	Win, success, better	0.09
Power	power	Superior, bully	0.02
Reward	reward	Take, prize, benefit	-0.11
Risk	risk	Danger, doubt	-0.13
Personal concerns			
Work	work	Job, majors	-0.07
Leisure	leisure	Cook, chat, movie	-0.21
Home	home	Kitchen, landlord	-0.10
Informal language	informal		-0.12
Nonfluencies	nonflu	Er, hm, umm	-0.15
Time orientation			
Past focus	focuspast	Ago, did, talked	-0.29
Present focus	focuspresent	Today, is, now	0.21
Future focus	focusfuture	May, will, soon	0.02

This table presents the twenty LIWC speech features (of 93 total) selected for use in this study. The sample includes 253 speech recordings from 44 adolescent subjects. The bolded values represent correlations between speech feature and concurrent A-LIFE depression scores with a $p < 0.1$.

Table 2

Example of affective processes LIWC feature within selected study speech samples.

LIWC category	Speech sample	LIWC analysis
Affective processes	“It’s been <u>fine</u> . The <u>play</u> has <i>stressed</i> me out. School has <i>stressed</i> me out, like being at school and <i>missing</i> school. Oh and my mom is <i>stressed</i> or <i>upset</i> . I’m <i>stressed</i> . Um uh my mom’s just uh <u>strong</u> and like any little thing will make her <i>upset</i> , so she’s very triggering, and that’s it.”	<ul style="list-style-type: none"> • Word Count: 54 words • 3 positive affect words (5.56 %); see underlined words • 7 negative affective (12.96 %); see italicized words • Affect score of 18.52 %

The speech feature of affective processes is exemplified by this speech sample. The LIWC analysis shows the total word count, the number and percentage positive affect words (underlined), the number and percentage of negative affect words (italicized), and the total percentage of affect words.