Received: 4 December 2018 | Revised: 14 February 2019 | Accepted: 8 March 2019

DOI: 10.1002/da.22890

RESEARCH ARTICLE

WILEY ADAA ANXIETY AND DEPRESSION ASSOCIATION OF AMERICA

Speech-based markers for posttraumatic stress disorder in US veterans

Charles R. Marmar[1,2] | Adam D. Brown[1,2,3] | Meng Qian[1,2] | Eugene Laska[1,2] | Carole Siegel[1,2] | Meng Li[1,2] | Duna Abu-Amara[1,2] | Andreas Tsiartas[4] | Colleen Richey[4] | Jennifer Smith[4] | Bruce Knoth[4] | Dimitra Vergyri[4]

[1]Department of Psychiatry, New York University School of Medicine, New York, New York

[2]Steven and Alexandra Cohen Veterans Center for the Study of Post-Traumatic Stress and Traumatic Brain Injury, New York, New York

[3]Department of Psychology, New School for Social Research, New York, New York

[4]Stanford Research Institute International, Menlo Park, California

Correspondence
Charles R. Marmar, M.D., Department of Psychiatry, New York University School of Medicine, 1 Park Avenue, New York, NY 10016.
Email: Charles.Marmar@nyulangone.org

Funding information
Telemedicine and Advanced Technology Research Center; U.S. Army Medical Research and Acquisition Activity (USAMRAA), Grant/Award Number: W81XWH-11-C-0004; Steven A. and Alexandra M. Cohen Foundation; Cohen Veterans Bioscience Inc. (CVB)

Abstract

Background: The diagnosis of posttraumatic stress disorder (PTSD) is usually based on clinical interviews or self-report measures. Both approaches are subject to under- and over-reporting of symptoms. An objective test is lacking. We have developed a classifier of PTSD based on objective speech-marker features that discriminate PTSD cases from controls.

Methods: Speech samples were obtained from warzone-exposed veterans, 52 cases with PTSD and 77 controls, assessed with the Clinician-Administered PTSD Scale. Individuals with major depressive disorder (MDD) were excluded. Audio recordings of clinical interviews were used to obtain 40,526 speech features which were input to a random forest (RF) algorithm.

Results: The selected RF used 18 speech features and the receiver operating characteristic curve had an area under the curve (AUC) of 0.954. At a probability of PTSD cut point of 0.423, Youden's index was 0.787, and overall correct classification rate was 89.1%. The probability of PTSD was higher for markers that indicated slower, more monotonous speech, less change in tonality, and less activation. Depression symptoms, alcohol use disorder, and TBI did not meet statistical tests to be considered confounders.

Conclusions: This study demonstrates that a speech-based algorithm can objectively differentiate PTSD cases from controls. The RF classifier had a high AUC. Further validation in an independent sample and appraisal of the classifier to identify those with MDD only compared with those with PTSD comorbid with MDD is required.

KEYWORDS

biomarkers, diagnostics, feature extraction, military, posttraumatic stress disorder, speech-based assessment, veterans

1 | INTRODUCTION

Posttraumatic stress disorder (PTSD) is frequently associated with functional impairment including relationship conflicts (Taft, Watkins, Stafford, Street, & Monson, 2011), reduced academic attainment (Bachrach & Read, 2012; Kessler, Foster, Saunders, & Stang, 1995),

Marmar and Brown should be have considered joint first authors.

Preliminary findings from this study were presented at the 16th annual conference of the International Speech Communication Association, Dresden, Germany, September 6–10, 2015.

substance abuse (Mills, Teesson, Ross, & Peters, 2006; Pietrzak, Goldstein, Southwick, & Grant, 2011), unemployment (Sripada et al., 2016), and adverse health outcomes (Boscarino, 2008; O'donovan, Slavich, Epel, & Neylan, 2013; Roberts et al., 2015; Zen, Whooley, Zhao, & Cohen, 2012). The ability to accurately screen for and diagnose PTSD, however, remains challenging (Shalev, Liberzon, & Marmar, 2017). There are numerous self-report screening tools (Sijbrandij et al., 2013) and several clinician-administered interview protocols (Blake et al., 1995; Foa & Tolin, 2000; Weathers et al., 2017). The gold-standard for diagnosing PTSD is the Clinician-Administered PTSD Scale (CAPS; Blake et al., 1995). The CAPS is a structured clinical interview for assessing the frequency and severity of PTSD symptoms and related functioning impairments. The CAPS has been shown to have 79% overall agreement with a clinician's diagnosis, with the sensitivity of 0.74 and specificity 0.84 (Hovens et al., 1994).

The assessment of PTSD with a structured interview is based in part on the subjective complaints of the patient and interpretations of the clinician. This process is subject to a number of biases that may distort the accuracy of the diagnosis, including cultural and racial biases (Snowden, 2003), distortions in memory (Donaldson, Corrigan, & Kohn, 2000; Ely, Graber, & Croskerry, 2011), or financial and social incentives (Hall & Hall, 2006). Additionally, because of stigma, patients vary in their willingness to candidly discuss traumatic experiences, symptoms and functioning. Moreover, the interview requires a lengthy visit to a clinician's office, which some patients may be unwilling or unable to do. For these reasons, there is an imperative to develop objective measures for screening and diagnosing psychiatric disorders (Kapur, Phillips, & Insel, 2012; Singh & Rose, 2009), including PTSD (Lehrner & Yehuda, 2014; Shalev et al., 2017).

Multiple studies have been initiated to identify biological markers for PTSD including alterations in neural structures and circuit functioning, genomics, neurochemistry, immune functioning, and psychophysiology (Lehrner & Yehuda, 2014; Shalev et al., 2017; Zoladz & Diamond, 2013). Despite these advances, problems in accuracy, cost, and patient burden preclude routine use in clinical practice.

There has been growing interest in speech-based techniques to screen for psychiatric disorders (Bedi et al., 2014, 2015; Grünerbl et al., 2015; Karam et al., 2014; Muaremi, Gravenhorst, Grünerbl, Arnrich, & Tröster, 2014; Osmani et al., 2015; Vanello et al., 2012). Speech is an attractive candidate, as it can be measured at low-cost, remotely, noninvasively, and naturalistically. Clinicians have long observed that individuals suffering from psychiatric disorders display changes in speech (Newman & Mather, 1938) and routinely use impressions of voice quality as an element of mental status examination, including "pressured" speech in bipolar disorder or "monotone," "lifeless," and "metallic" speech in depression (Hall, Harrigan, & Rosenthal, 1995; Moses, 1954; Sobin & Sackeim, 1997). More recently, automated techniques to analyze speech have been able to classify mood disorders on a number of speech features. For example, combining prosodic, voice quality, spectral, and glottal features for automated speech classification has shown encouraging sensitivity and specificity (van den Broek, van der Sluis & Dijkstra, 2010).

Less is known about speech alterations in PTSD. Van den Broek, Van der Sluis, and Dijkstra (2010) asked individuals with PTSD to generate two affective narratives and found that 65 parameters of speech accounted for 69–83% of the variance of stress symptoms. Scherer et al. found that in response to positive, negative, and neutral interview prompts, those with PTSD exhibited more tense voice features (Scherer, Stratou, Gratch, & Morency, 2013) and decreased vowel space (Scherer, Lucas, Gratch, Rizzo, & Morency, 2016). Recent work applying multiview learning algorithms demonstrated that diagnostic classification of PTSD increased by 20–37% using two speech classifiers (Zhuang, Rozgić, Crystal, & Marx, 2014). Although promising, these findings are limited due to reliance on self-report measures rather than validated interviews to classify PTSD (Scherer et al., 2013, 2016), samples with major depressive disorder (MDD) comorbidity (Scherer et al., 2013, 2016), limited use of control groups (van den Broek, van der Sluis & Dijkstra, 2010), and small samples (Scherer et al., 2013, 2016; van den Broek, van der Sluis & Dijkstra, 2010).

This is the first study to identify features of speech that differentiate PTSD cases from controls in an age- and gender-matched sample of veterans excluding current MDD.

2 | METHODS

2.1 | Participants

Participants included 129 American warzone-exposed male Iraq and Afghanistan veterans who gave written informed consent. All procedures were approved by the Institutional Review Board of NYU Langone School of Medicine and conform to the US Federal Policy for the Protection of Human Rights. Participants were assessed for PTSD with the Clinician Administered PTSD Scale (CAPS-IV) by a clinical psychologist. Participants in the PTSD group met diagnostic criteria for PTSD based on DSM IV-TR criteria (Blake et al., 1990). Controls were age- and gender-matched warzone-exposed veterans who did not meet criteria for current or lifetime PTSD.

Participants were excluded from the study if they met DSM 5 criteria, assessed by the Structured Clinical Interview for DSM Diagnosis (SCID-5), severe drug use in the past 6 months, lifetime history of any psychiatric disorder with psychotic features, bipolar I & II disorder, current MDD, depression due to a general medical condition (GMC), current exposure to recurrent trauma or exposure to a traumatic event within the past month, prominent suicidal ideation, homicidal ideation, suicide attempt in the past 3 months, history of open-head injury, illness affecting central nervous system (CNS) functioning, cardiovascular disease, major medical illness, and starting psychotropic medications in the past month.

2.2 | Procedure

2.2.1 | Speech feature extraction

The audio of each CAPS interview was recorded in two channels, using separate microphones for the interviewer and participant. A rich set of speech features was extracted from the participant's recording using the following steps.

2.2.2 | Audio quality control

This step was manual, targeting the selection of only good audio samples (clear and audible speech in the signal) to avoid noise in the feature extraction process. During this step the participant's audio channel was also manually marked.

2.2.3 | Audio segmentation

This step identified the participant's speech regions, excluding the interviewer who was often audible in the participant microphone channel. Very short duration (e.g., "yes/no") participant segments were also removed, resulting in 1–120 min of clean speech per participant (mean = 35 min/speaker). This step could also be done manually, but due to cost and time constraints we applied SRI's automatic Voice Activity Detector (VAD; open source alternatives can also be used, e.g., https://chromium.googlesource.com/external/webrtc/ + /master/common_audio/vad/). VAD was run on the clinician and participant audio channels independently to mark the locations of speech for both speakers. Only the participant channel speech regions with higher VAD score than the corresponding clinician channel segments were retained, avoiding segments that included interviewer's voice. In the following, we refer to these automatically identified participant speech regions, separated by long pauses or speech from the interviewer, as speech "spurts."

2.2.4 | Extraction of frame level features

A frame is a short sliding window of speech, typically 5–25 ms, depending on feature type. The frame-level features included: spectral (i.e., Mel-Frequency Cepstral Coefficients [MFCCs]), linear predictive coding (LPC), noise-robust spectral (i.e., DOCC and RASTA), prosodic (chroma features, pitch, voicing, and correlation), time-based (zero crossing, RMS energy, and L1-norm), spectro-temporal (LTSV, MFCC, and RASTA derivatives), articulatory, temporal, and machine-learning-based (autoencoders learned from prior speaker databases). The features were extracted using SRI's speech feature extraction tools, but there are also open source alternatives that can be used for the same feature types (e.g., https://www.audeering.com/opensmile/).

2.2.5 | Computation of spurt-level features

These were computed based on frame-level features for every spurt. They included: (a) statistics: mean, variance, kurtosis skewness, variation from mean, percentiles, range, and slope, (b) locational information: absolute and relative distances from the beginning of the spurt for the occurrence of important feature values (min, max, 5% of max, 50% of max, 95% of max) and (c) durational information: distances between the occurrences of important feature values, for example, the distance between reaching 5% and 50% of the feature max value within the spurt.

2.2.6 | Computation of speaker-level features

The final feature vector was extracted by taking statistics of the spurt level features for each speaker: mean variance, kurtosis, skewness, variation from mean, various percentiles, interquartile range, and slope.

These features aim to capture the nuances, variability, and behavior, both short-term and long-term, of a rich set of low-level speech features over the entire session focusing only on the patient speech segments of the conversation. A total of 40,526 features were computed at the speaker level and were used for the feature selection and model building.

2.3 | Statistical analysis

2.3.1 | Comparing the two groups on demographic variables

For categorical variables, a χ^2 (or Fisher's exact test when at least one cell count had five or less individuals) and for continuous variables, Wilcoxon's rank sum tests were used.

2.3.2 | The random forest (RF) probabilistic classifier

A RF algorithm was used to build a classifier function using speech markers to predict PTSD. It is an algorithm (Breiman, Friedman, Olshen, & Stone, 1984; Malley, Kruppa, Dasgupta, Malley, & Ziegler, 2012) based on multiple classification and regression trees (CART; Strobl, Malley, & Tutz, 2009) yielding a probability estimate of membership in a target prediction class based on marker values. CART grows a decision tree whose hierarchical nodes are each based on a cut-point split of a predictor found by an exhaustive search to minimize misclassification error. The process continues recursively until a tree is grown with nodes that contain members from only one group. This tree is pruned to a set of nodes for which little is gained from further splits in improving misclassification error. An estimate of the probability of membership in the target group of an individual in a terminal node is given by the fraction of members in the target group who are in the node. RF makes use of an ensemble of CART decision trees for prediction which acts to decrease the variance of the predictions and the inherent potential of over-fitting of a single decision tree. Bootstrap samples of subjects can be used to grow a RF of trees. Data on the "out-of-bag" (OOB) subjects in each sample, consisting of approximately one-third of the full sample whose data were not used to grow the particular tree, are used to obtain predictions of target class membership. Features of the OOB subjects are scored and the estimate of the probability of being in the target class is the fraction of the target class in the terminal node into which they fall. The average of these estimates over the trees grown is the RF estimate of the probability. These are then used to generate a receiver operator curve and its area under the curve (AUC).

The importance of a predictor is assessed by randomly permuting its value in the OOB sample and comparing the differences in predictive performance (AUCs) between the nonpermuted and permuted samples. The AUCs are averaged across the entire forest and ranked on the decrease in AUCs (Breiman et al., 1984). "Shaving" is a method for reducing the number of predictors based on variable importance. The variable of least importance is shaved off first and a new RF is obtained. The procedure is repeated until all variables have been shaved. The shaved RF with a parsimonious mix of a small number of features and a large AUC is chosen.

In this study, based on 20,000 bootstrap samples, a RF based on the 40,759 voice markers was grown and the shaving step began starting with the 500 variables with the highest importance rankings.

2.3.3 | Testing for confounding

We tested for the possibility that the findings of the relationship between voice markers and PTSD are confounded by the presence of comorbidities of traumatic brain injury (TBI), alcohol use disorder (AUD), and symptoms of depression. Participants who met criteria for PTSD comorbid with MDD had been excluded. Residual symptoms of depression were measured by the Beck Depression Inventory-II (BDI-II). A variable was considered to be a confounder if two null hypotheses related to the prediction of PTSD were rejected (Pearl, 2009). The null hypotheses to be rejected are (a) that the potential confounder is not associated with the predictive voice markers and (b) that the probability of being a PTSD case is not different when including the confounder in the model from predicting with the voice markers alone. The confounder hypothesis tests were run separately for TBI, AUD, the individual symptoms of depression, and total BDI-II score. For tests of the first confounder hypothesis, χ^2 tests were run on contingency tables of confounder by voice marker. For tests of the second hypothesis, estimates of the probability of PTSD obtained from the final RF with and without the inclusion of the candidate confounder were obtained and compared using Wilcoxon's rank sum test. If the latter test was statistically significant, we also required that the difference in AUCs be >0.05.

3 | RESULTS

3.1 | Demographics

The PTSD cases and controls did not differ significantly by age, ethnicity, educational attainment, number of warzone deployments or current cannabis, cocaine, hallucinogen, opioid, or stimulant use (Table 1). The PTSD group had significantly higher total BDI scores, TBI exposure levels, and current rates of AUD.

3.2 | Properties of the RF

The final shaved RF selected was based on 18 voice markers with an AUC = 0.954. At a PTSD probability cut point of 0.423, Youden's index, defined as the sum of sensitivity + specificity − 1, was 0.904 + 0.883 − 1 = 0.787 with an overall correct classification rate of 89.1%.

TABLE 1 Participant demographics

Variables	PTSD+ (*N* = 52) *N* (%) or mean (*SD*)	PTSD− (*N* = 77) *N* (%) or mean (*SD*)
Age (years)	31.92 (5.97)	32.47 (7.22)
Number of deployments	1.73 (1.01)	1.79 (1.09)
Race		
Asian	2 (3.9%)	5 (6.5%)
Black/African American	9 (17.3%)	6 (7.8%)
White/Caucasian	29 (55.8%)	46 (59.7%)
Hispanic/Latino	11 (21.2%)	14 (18.2%)
Other	1 (1.9%)	6 (7.8%)
Education		
Up to 12th grade	1 (1.9%)	0 (0.0%)
High school/GED	18 (34.6%)	17 (22.1%)
2 years college/Associate's degree	14 (26.9%)	14 (18.2%)
4 years college/Bachelor's degree	12 (23.1%)	32 (41.6%)
Master's degree	7 (13.5%)	14 (18.2%)
TBI exposure		
Yes	19 (36.5%)	5 (6.5%)*
Current alcohol use		
Yes	14 (26.9%)	5 (6.5%)*
Current cannabis use		
Yes	2 (3.9%)	0 (0%)
Current cocaine use		
Yes	0 (0%)	0 (0%)
Current hallucinogen use		
Yes	1 (1.9%)	0 (0%)
Current stimulant use		
Yes	0 (0%)	0 (0%)
Current opioid use		
Yes	0 (0%)	0 (0%)
BDI total score[a]	12.54 (8.65)	3.59 (4.15)*

Note. BDI: beck depression inventory; PTSD: posttraumatic stress disorder.

*Significant $p < 0.05$.

[a] BDI total score is the sum of all 21 BDI items.

3.3 | Voice marker features in the RF

Table 2 lists the 18 features used to build the selected model. Among individuals with PTSD, Feature 3 reflects speech segments containing articulators that move more slowly than in controls or contain long extended vowels, including hesitations (e.g., "uh...."). In addition, features 1, 2, 4, 5, 11, 15, and 17 contain speech features that were more monotonous in PTSD cases than in controls. Additionally,

TABLE 2 Feature description for top (ordered by variable importance score) 18 features

Feature #	Quality of speech	Description of feature computation
1	More monotonous speech (less varying tonality)	For each spurt we computed the relative time distance between the occurrence of the low (5%) and median (50%) values for a specific spectral feature (first chroma FFT coefficient), representing variability in certain speech frequencies. Then we extracted the lowest value across the speaker spurts.
2	Monotonous speech segments	For each spurt we computed the relative time distance between the maximum and the minimum values for a specific spectral feature (second LMFCC coefficient), representing variability in certain speech frequencies. Then we extracted the lowest value across the speaker spurts.
3	Occurrences of slow speech production	For each speech spurt we estimated the average time it took for the tongue to move from the minimum to the maximum point. Then we extracted the highest value (slowest changing spurt) across each speaker's speech.
4	More monotonous speech (less varying tonality)	For each spurt we computed the relative time distance between the occurrence of the maximum and the median value for a tonal frequency, representing the tonal variability on a certain frequency. Then we found the average value across all spurts.
5	Less bursty (more monotonous) voice	For each spurt we computed the kurtosis value for a specific spectral feature (third Chroma filter) detecting existence of anomalies/outliers in the distribution of certain speech frequencies. Then we extracted the skewness of this value across the speaker spurts. This measured whether there were outliers (bursts) in speech tonality or whether it was mostly within expected ranges during the session.
6	Flatter speech	For each spurt we computed the normalized variance of a specific spectral feature (11th LMFCC coefficient). Then we computed the kurtosis (consistency) across each speaker's spurts.
7	Less animated speech	For each spurt we computed the skewness for a specific spectral feature (11th LMFCC coefficient) which represented the symmetry of the distribution (found if there was an outlier values). Then we extracted the lowest value across the speaker spurts. It examined the least varying spurt, which may have contained single vowel sounds.
8	Speech segments with very low activation	For each spurt we computed the relative time distance between the occurrence of the minimum and the maximum value for a specific spectral feature (11th LMFCC coefficient) representing variability in certain speech frequencies. Then we extract the lowest value across the speaker spurts
9	Flatter speech in terms of energy variation	For each spurt we computed the highest tonal energy for a certain frequency range (chroma FFT ninth coefficient). Then we computed the variability of that energy across all spurts.
10	Flat tone in speech	For each spurt we computed the highest tonal energy for a certain frequency range (chroma FFT th coefficient). As for feature 9 but take the 95th% across all spurts.
11	More monotonous speech	For each spurt we computed a tonal frequency (zeroth chroma FFT coefficient) with the highest value across all spurts.
12	Flatter speech in terms of energy variation	Similar to feature 6, but examined the 10th LMFCC coefficient.
13	Less activated speech	For each spurt we computed the skewness of a specific spectrogram frequency range and then we find the minimum across all spurts. It measured the flatness of the spectrogram for that frequency range.
14	Slow speech production or long hesitations	For each spurt we estimated the kurtosis of the position of an articulator. Then we extracted the highest value (flattest spurt) across the speaker spurts. Similar to feature 3, it was examining the most consistently articulated spurt
15	More monotonous speech (less varying tonality)	For each spurt we get the range of values for a specific spectral features (fourth Chroma filter) and compute the deviation of the range across all spurts. This indicates how variable is the range of tonality across spurts
16	Flatter speech in terms of energy variation	For each spurt we computed the relative time distance between the occurrence of the low (5%) and median (50%) value for specific spectral frequencies (low frequency range), representing how fast the energy changed within that range. Then we extracted the skewness across the speaker spurts, which showed whether that energy changed in a consistent manner across all spurts

(Continues)

TABLE 2 (Continued)

Feature #	Quality of speech	Description of feature computation
17	More monotonous speech in energy	For each spurt we compute the 5% value for a certain spectral feature (19th Rasta coeff.) and compute the deviation of this value across spurts. In captures energy variability in a certain frequency range
18	Description of speech quality could not be made	For each spurt we estimated the kurtosis of the position of an articulator. Then we extracted the highest value (flattest spurt) across the speaker spurts. Similar to feature 1, it was examining the most consistently articulated spurt

features 6, 9, 10, 12, and 16 revealed that individuals with PTSD were more likely to generate flat speech. Moreover, features 8 and 13 contained speech features indicating less speech activation among cases. Table 3 displays means, standard deviations (SDs), and medians of each feature and results of the Wilcoxon test comparing the distributions of speech markers between groups. All but feature 18 significantly differed between cases and controls.

3.4 | Confounders

BDI symptoms, TBI, and AUD failed to meet statistical criteria required to confirm that they are confounders. For the tests of independence of potential confounders with each marker, TBI and AUD were correlated only with feature 12. BDI symptoms were each individually correlated with at most two markers. For the second statistical test, Table 4 shows the results of comparisons of the probabilities of PTSD and the AUCs of the RFs determined with and without inclusion of the confounders for BDI total score, TBI, and AUD as predictors. While estimates of the probability of PTSD differed between the two models, AUCs were not improved by including either TBI or alcohol use. In addition, individual BDI symptom scores did not significantly increase AUCs and the BDI total score improved AUC by only 1%. Appendix I contains results of the complete confounder analyses including tests for each BDI symptom.

4 | DISCUSSION

This study demonstrated that by using speech-based techniques, male Iraq and Afghanistan veterans with PTSD could be distinguished from warzone-exposed veterans without PTSD, neither of whom had MDD. The findings suggest that by combining frame-level short- and longer-duration prosodic features, high accuracy, sensitivity, and specificity for classifying PTSD can be achieved. The classifier assigns

TABLE 3 Summary statistics of 18 voice markers for the PTSD− and the PTSD+ groups

	PTSD−			PTSD+			
Variables	Mean	Median	*SD*	Mean	Median	*SD*	Wilcoxon's test
Var1	−0.964	−0.973	0.030	−0.982	−0.987	0.024	*
var2	−0.937	−0.942	0.035	−0.965	−0.970	0.022	*
var3	0.936	0.945	0.053	0.967	0.972	0.021	*
var4	−0.065	−0.081	0.084	−0.039	−0.059	0.095	*
var5	409.400	240.187	557.587	1026.070	630.206	1154.760	*
var6	2.682	2.139	2.498	2.862	2.744	0.757	*
var7	−0.959	−0.967	0.038	−0.980	−0.983	0.014	*
var8	0.279	0.269	0.048	0.249	0.250	0.039	*
var9	−1.430	−1.364	0.336	−1.763	−1.657	0.632	*
var10	0.004	0.003	0.002	0.003	0.003	0.001	*
var11	−1.810	−1.883	0.463	−2.316	−2.271	1.120	*
var12	0.929	0.945	0.074	0.940	0.970	0.196	*
var13	355.616	208.173	364.526	897.390	605.526	769.039	*
var14	12.838	12.131	4.279	17.006	15.581	5.937	*
var15	0.00024	0.00018	0.00017	0.00018	0.00015	0.00019	*
var16	0.035	0.164	1.617	−0.131	−0.208	1.883	*
var17	0.0040	0.0037	0.0015	0.0032	0.0031	0.0010	*
var18	0.170	0.169	0.012	0.171	0.169	0.006	

Note. PTSD: posttraumatic stress disorder; *SD*: standard deviation.
$^*p < 0.05$.

TABLE 4 Random forest results: Wilcoxon's test comparing probabilities of PTSD and AUCs with and without confounders—TBI, alcohol use, and BDI total score

Variable	Wilcoxon's test	AUC of model with 18 markers	AUC of model with 18 markers plus confounder
TBI	*	0.954	0.954
Alcohol_use	*	0.954	0.954
BDI_total_score	*	0.946	0.957

Note. AUC: area under the curve; BDI: beck depression inventory; PTSD: posttraumatic stress disorder; TBI: traumatic brain injury.
*$p < 0.05$.

higher probabilities of PTSD to those with features indicating speech that is slower, more monotonous, and less change in tonality and activation.

Although the biological mechanisms underlying the link between these speech features and PTSD were not examined in this study, previous work has documented that changes to the automatic nervous system cause disturbances in similar speech-based features, such as muscle tension (Scherer, 1986) and respiratory rate (Kreibig, 2010). Additionally, the neurotransmitter gamma-amino butyric acid (GABA), which has been linked with a vulnerability to both depression (Croarkin, Levinson, & Daskalakis, 2011) and suicidality (Poulter et al., 2008), has also been associated with changes in muscle tonality (Croarkin et al., 2011). Importantly, GABA has been identified as a promising neural marker of vulnerability and resilience to PTSD, as well as a therapeutic target (Faye, McGowan, Denny, & David, 2018; Kelmendi et al., 2016).

Furthermore, changes in muscle tension alter vocal tract dynamics and constrain articulatory movement. A speaker's level of depression has been shown to affect prosodic and source features relevant to articulation (Moore, Clements, Peifer & Weisser, 2008; Mundt, Vogel, Feltner, & Lenderking, 2012; Quatieri & Malyska, 2012; Scherer et al., 2013; Trevino, Quatieri, & Malyska, 2011) and has been correlated with reduced tonal range (Breznitz, 1992; Darby, Simmons, & Berger, 1984; Flint, Black, Campbell-Taylor, Gailey, & Levinton, 1993). Findings from the current study indicate that a reduction in tonal range may also be associated with PTSD, even among individuals who do not meet criteria for MDD. The reduction in tonality observed in PTSD is consistent with previous studies showing decreased formant frequencies in depressed individuals (Mundt, Snyder, Cannizzaro, Chappie, & Geralts, 2007). These findings may reflect that tonality (e.g., F0 frequency) is influenced by factors such as current mood (Ellgring & Scherer, 1996), level of agitation and anxiety (Alpert, Pouget, & Silva, 2001; Tolkmitt, Helfrich, Standke, & Scherer, 1982), and personality traits (Yang, Fairbairn, & Cohn, 2013). Although the exact processes contributing to the observed reduction in tonality in this study were not tested, future experimental studies would allow for a more specific understanding.

Taken together, these data offer strong preliminary evidence that speech features can serve as an objective probability classifier for PTSD. Compared to the more extensive literature linking speech-based features and mood disorders (Cummins et al., 2015), there has been a paucity of research examining speech in PTSD. The few published studies relied on small sample sizes, assessed PTSD with self-report measures, and had high levels of comorbid MDD, making it difficult to determine whether those features were associated with PTSD or related psychopathologies.

This is the first study to use a structured clinical interview, the CAPS 5, both for classifying cases and controls and for the collection of speech segments for vocal analysis. The ability to use data collected naturalistically suggests that clinicians may be able to employ speech-based analyses to aid in the diagnostic process from information routinely collected by clinicians. In contrast, the CAPS interview may have been more stressful for those with PTSD, compared to controls. It is unclear whether these differences are only found under conditions of stress or if they would be found in speech segments generated from less affectively charged content.

There were a number of limitations in the study. While we have conducted extensive internal cross validation, classifier endorsement requires a newly recruited external validation sample. We are confident that TBI and AUD did not confound voice marker findings in this study because there are a substantial number of subjects with these disorders in the sample, yielding sufficient power for the confounder analyses. Nevertheless, larger sample sizes in future studies would increase confidence in these findings.

Previous work suggested that similar alterations in speech are associated with affective dysregulation (Breiman, 2001). For example, “monotony” and “dullness” have long been associated with a depressed or sad voice. Kraepelin (1921) described speech quality of depressed patients as “low voice, slowly, hesitatingly, monotonously, sometimes stuttering, whispering.” The question of whether the panel predicts depression rather than PTSD must be considered. To minimize this possibility, participants with MDD were excluded from both groups. Further, for symptomatology not meeting criteria for MDD, we tested the BDI symptoms and did not find them to be confounders. Clarification of the value of the classifier in clinical settings requires studies of persons with the diagnosis of MDD without PTSD and those with comorbid MDD and PTSD.

Given these limitations, we believe that our panel of voice markers represents a rich, multidimensional set of features which with further validation holds promise for developing an objective, low cost, noninvasive, and, given the ubiquity of smart phones, widely accessible tool for assessing PTSD in veteran, military, and civilian contexts.

ACKNOWLEDGEMENTS

This study was supported by the US Army Medical Research & Acquisition Activity (USAMRAA), the Telemedicine and Advanced Technology Research Center (TATRC), W81XWH-11-C-0004, and grants from the Steven A. and Alexandra M. Cohen Foundation, Inc.

and Cohen Veterans Bioscience Inc. (CVB) to NYU Langone School of Medicine.

ORCID

Charles R. Marmar http://orcid.org/0000-0001-8427-5607
Adam D. Brown http://orcid.org/0000-0002-6151-5257

REFERENCES

Alpert, M., Pouget, E. R., & Silva, R. R. (2001). Reflections of depression in acoustic measures of the patient's speech. *Journal of Affective Disorders*, *66*(1), 59–69. https://doi.org/10.1016/S0165-0327(00)00335-9

Bachrach, R. L., & Read, J. P. (2012). The role of posttraumatic stress and problem alcohol involvement in university academic performance. *Journal of Clinical Psychology*, *68*(7), 843–859. https://doi.org/10.1002/jclp.21874

Bedi, G., Carrillo, F., Cecchi, G. A., Slezak, D. F., Sigman, M., Mota, N. B., & Corcoran, C. M. (2015). Automated analysis of free speech predicts psychosis onset in high-risk youths. *NPJ Schizophrenia*, *1*, 15030. https://doi.org/10.1038/npjschz.2015.30

Bedi, G., Cecchi, G. A., Slezak, D. F., Carrillo, F., Sigman, M., & De Wit, H. (2014). A window into the intoxicated mind? Speech as an index of psychoactive drug effects. *Neuropsychopharmacology*, *39*(10), 2340–2348. https://doi.org/10.1038/npp.2014.80

Blake, D. D., Weathers, F., Nagy, L. M., Kaloupek, D. G., Klauminzer, G., Charney, D. S., & Keane, T. M. (1990). A clinician rating scale for assessing current and lifetime PTSD: The CAPS-1. *The Behavior Therapist*, *13*, 187–188.

Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Gusman, F. D., Charney, D. S., & Keane, T. M. (1995). The development of a clinician-administered PTSD scale. *Journal of Traumatic Stress*, *8*(1), 75–90. https://doi.org/10.1007/BF02105408

Boscarino, J. A. (2008). A prospective study of PTSD and early-age heart disease mortality among Vietnam veterans: Implications for surveillance and prevention. *Psychosomatic Medicine*, *70*(6), 668–676. https://doi.org/10.1097/PSY.0b013e31817bccaf

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees, *The Wadsworth Statistics/Probability Series*. Monterey, CA: Wadsworth and Brooks.

Breznitz, Z. (1992). Verbal indicators of depression. *The Journal of General Psychology*, *119*(4), 351–363. https://doi.org/10.1080/00221309.1992.9921178

van den Broek, E. L., van der Sluis, F., & Dijkstra, T. (2010). Telling the story and re-living the past: How speech analysis can reveal emotions in post-traumatic stress disorder (PTSD) patients. *Sensing Emotions*, 153–180. https://doi.org/10.1007/978-90-481-3258-4_10

Croarkin, P. E., Levinson, A. J., & Daskalakis, Z. J. (2011). Evidence for GABAergic inhibitory deficits in major depressive disorder. *Neuroscience & Biobehavioral Reviews*, *35*(3), 818–825. https://doi.org/10.1016/j.neubiorev.2010.10.002

Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication*, *71*, 10–49. https://doi.org/10.1016/j.specom.2015.03.004

Darby, J. K., Simmons, N., & Berger, P. A. (1984). Speech and voice parameters of depression: A pilot study. *Journal of Communication Disorders*, *17*(2), 75–85. https://doi.org/10.1016/0021-9924(84)90013-3

Donaldson, M. S., Corrigan, J. M., & Kohn, L. T. (Eds.) (2000). To err is human:Building a safer health system (vol. 6). Washington, DC: National Academies Press.

Ellgring, H., & Scherer, K. R. (1996). Vocal indicators of mood change in depression. *Journal of Nonverbal Behavior*, *20*(2), 83–110. https://doi.org/10.1007/BF02253071

Ely, J. W., Graber, M. L., & Croskerry, P. (2011). Checklists to reduce diagnostic errors. *Academic Medicine*, *86*(3), 307–313. https://doi.org/10.1097/ACM.0b013e31820824cd

Faye, C., McGowan, J. C., Denny, C. A., & David, D. J. (2018). Neurobiological mechanisms of stress resilience and implications for the aged population. *Current Neuropharmacology*, *16*(3), 234–270. https://doi.org/10.2174/1570159X15666170818095105

Flint, A. J., Black, S. E., Campbell-Taylor, I., Gailey, G. F., & Levinton, C. (1993). Abnormal speech articulation, psychomotor retardation, and subcortical dysfunction in major depression. *Journal of Psychiatric Research*, *27*(3), 309–319. https://doi.org/10.1016/0022-3956(93)90041-Y

Foa, E. B., & Tolin, D. F. (2000). Comparison of the PTSD symptom scale–interview version and the clinician-administered PTSD scale. *Journal of Traumatic Stress: Official Publication of the International Society for Traumatic Stress Studies*, *13*(2), 181–191. https://doi.org/10.1023/A:1007781909213

Grünerbl, A., Muaremi, A., Osmani, V., Bahle, G., Oehler, S., Tröster, G., & Lukowicz, P. (2015). Smartphone-based recognition of states and state changes in bipolar disorder patients. *IEEE Journal of Biomedical and Health Informatics*, *19*(1), 140–148. https://doi.org/10.1109/JBHI.2014.2343154

Hall, J. A., Harrigan, J. A., & Rosenthal, R. (1995). Nonverbal behavior in clinician—patient interaction. *Applied and Preventive Psychology*, *4*(1), 21–37. https://doi.org/10.1016/S0962-1849(05)80049-6

Hall, R. C., & Hall, R. C. (2006). Malingering of PTSD: Forensic and diagnostic considerations, characteristics of malingerers and clinical presentations. *General Hospital Psychiatry*, *28*(6), 525–535. https://doi.org/10.1016/j.genhosppsych.2006.08.011

Hovens, J. E., Van der Ploeg, H. M., Klaarenbeek, M. T. A., Bramsen, I., Schreuder, J. N., & Rivero, V. V. (1994). The assessment of posttraumatic stress disorder: With the Clinician Administered PTSD Scale: Dutch results. *Journal of Clinical Psychology*, *50*(3), 325–340. https://doi.org/10.1002/1097-4679(199405)50:3<325::AID-JCLP2270500304>3.0.CO;2-M

Kapur, S., Phillips, A. G., & Insel, T. R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Molecular Psychiatry*, *17*(12), 1174–1179. https://doi.org/10.1038/mp.2012.105

Karam, Z. N., Provost, E. M., Singh, S., Montgomery, J., Archer, C., Harrington, G., & Mcinnis, M. G. (2014). Ecologically valid long-term mood monitoring of individuals with bipolar disorder using speech. *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference*, 4858–4862. https://doi.org/10.1109/ICASSP.2014.6854525.

Kelmendi, B., Adams, T. G., Yarnell, S., Southwick, S., Abdallah, C. G., & Krystal, J. H. (2016). PTSD: From neurobiology to pharmacological treatments. *European Journal of Psychotraumatology*, *7*(1), 31858. https://doi.org/10.3402/ejpt.v7.31858

Kessler, R. C., Foster, C. L., Saunders, W. B., & Stang, P. E. (1995). Social consequences of psychiatric disorders, I: Educational attainment. *American Journal of Psychiatry*, *152*(7), 1026–1032. https://doi.org/10.1176/ajp.152.7.1026

Kraepelin, E. (1921). Manic depressive insanity and paranoia. *The Journal of Nervous and Mental Disease*, *53*(4), 350.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84*(3), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010

Lehrner, A., & Yehuda, R. (2014). Biomarkers of PTSD: Military applications and considerations. *European Journal of Psychotraumatology*, *5*, 10.3402/ejpt.v5.23797. https://doi.org/10.3402/ejpt.v5.23797

Malley, J. D., Kruppa, J., Dasgupta, A., Malley, K. G., & Ziegler, A. (2012). Probability machines. *Methods of Information in Medicine*, *51*(1), 74–81. https://doi.org/10.3414/ME00-01-0052

Mills, K. L., Teesson, M., Ross, J., & Peters, L. (2006). Trauma, PTSD, and substance use disorders: Findings from the Australian National

Survey of Mental Health and Well-Being. *American Journal of Psychiatry*, *163*(4), 652–658. https://doi.org/10.1176/ajp.2006.163.4.652

Moore, E., II, Clements, M. A., Peifer, J. W., & Weisser, L. (2008). Critical analysis of the impact of glottal features in the classification of clinical depression in speech. *IEEE Transactions on Biomedical Engineering*, *55*(1), 96–107. https://doi.org/10.1109/TBME.2007.900562

Moses, P. J. (1954). *The Voice of Neurosis*. New York, NY: Grune & Stratton.

Muaremi, A., Gravenhorst, F., Grünerbl, A., Arnrich, B., & Tröster, G. (2014). Assessing bipolar episodes using speech cues derived fromphone calls. *International Symposium on Pervasive Computing Paradigmsfor Mental Health*, 103–114. https://doi.org/10.1007/978-3-319-11564-1_11.

Mundt, J. C., Snyder, P. J., Cannizzaro, M. S., Chappie, K., & Geralts, D. S. (2007). Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *Journal of Neurolinguistics*, *20*(1), 50–64. https://doi.org/10.1016/j.jneuroling.2006.04.001

Mundt, J. C., Vogel, A. P., Feltner, D. E., & Lenderking, W. R. (2012). Vocal acoustic biomarkers of depression severity and treatment response. *Biological Psychiatry*, *72*(7), 580–587. https://doi.org/10.1016/j.biopsych.2012.03.015

Newman, S., & Mather, V. G. (1938). Analysis of spoken language of patients with affective disorders. *American Journal of Psychiatry*, *94*(4), 913–942. https://doi.org/10.1176/ajp.94.4.913

O'donovan, A., Slavich, G. M., Epel, E. S., & Neylan, T. C. (2013). Exaggerated neurobiological sensitivity to threat as a mechanism linking anxiety with increased risk for diseases of aging. *Neuroscience & Biobehavioral Reviews*, *37*(1), 96–108. https://doi.org/10.1016/j.neubiorev.2012.10.013

Osmani, V., Gruenerbl, A., Bahle, G., Haring, C., Lukowicz, P., & Mayora, O. (2015). Smartphones in mental health: Detecting depressive and manic episodes. *IEEE Pervasive Computing*, *14*(3), 10–13. https://doi.org/10.1109/MPRV.2015.54

Pearl, J. (2009). Causality: Models, *Reasoning, and Inference* (second ed.). Cambridge, UK: Cambridge University Press.

Pietrzak, R. H., Goldstein, R. B., Southwick, S. M., & Grant, B. F. (2011). Prevalence and Axis I comorbidity of full and partial posttraumatic stress disorder in the United States: Results from Wave 2 of the National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Anxiety Disorders*, *25*(3), 456–465. https://doi.org/10.1016/j.janxdis.2010.11.010

Poulter, M. O., Du, L., Weaver, I. C., Palkovits, M., Faludi, G., Merali, Z., & Anisman, H. (2008). GABAA receptor promoter hypermethylation in suicide brain: Implications for the involvement of epigenetic processes. *Biological Psychiatry*, *64*(8), 645–652. https://doi.org/10.1016/j.biopsych.2008.05.028

Quatieri, T. F., & Malyska, N. (2012). Vocal-source biomarkers for depression: A link to psychomotor activity. *Interspeech*, *2012*, 1059–1062.

Roberts, A. L., Agnew-Blais, J. C., Spiegelman, D., Kubzansky, L. D., Mason, S. M., Galea, S., & Koenen, K. C. (2015). Posttraumatic stress disorder and incidence of type 2 diabetes mellitus in a sample of women: A 22-year longitudinal study. *JAMA Psychiatry*, *72*(3), 203–210. https://doi.org/10.1001/jamapsychiatry.2014.2632

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, *99*(2), 143–165. https://doi.org/10.1037/0033-2909.99.2.143

Scherer, S., Lucas, G. M., Gratch, J., Rizzo, A. S., & Morency, L. P. (2016). Self-reported symptoms of depression and PTSD are associated with reduced vowel space in screening interviews. *IEEE Transactions on Affective Computing*, *7*(1), 59–73. https://doi.org/10.1109/TAFFC.2015.2440264

Scherer, S., Stratou, G., Gratch, J., & Morency, L. P. (2013). Investigating voice quality as a speaker-independent indicator of depression and PTSD. *Annual Conference of the International Speech Communication Association* (Interspeech), 847–851.

Scherer, S., Stratou, G., Mahmoud, M., Boberg, J., Gratch, J., Rizzo, A., & Morency, L. P. (2013). Automatic behavior descriptors for psychological disorder analysis. *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 1–8. https://doi.org/10.1109/FG.2013.6553789.

Shalev, A., Liberzon, I., & Marmar, C. (2017). Post-traumatic stress disorder. *New England Journal of Medicine*, *376*(25), 2459–2469. https://doi.org/10.1056/NEJMra1612499

Sijbrandij, M., Reitsma, J. B., Roberts, N. P., Engelhard, I. M., Olff, M., Sonneveld, L. P., & Bisson, J. I. (2013). Self-report screening instruments for post-traumatic stress disorder (PTSD) in survivors of traumatic experiences (protocol). *Cochrane Database of Systematic Reviews*, *2013*(6), 1–15. https://doi.org/10.1002/14651858.CD010575

Singh, I., & Rose, N. (2009). Biomarkers in psychiatry. *Nature*, *460*(7252), 202–207. https://doi.org/10.1038/460202a

Snowden, L. R. (2003). Bias in mental health assessment and intervention: Theory and evidence. *American Journal of Public Health*, *93*(2), 239–243. https://doi.org/10.2105/AJPH.93.2.239

Sobin, C., & Sackeim, H. A. (1997). Psychomotor symptoms of depression. *American Journal of Psychiatry*, *154*(1), 4–17. https://doi.org/10.1176/ajp.154.1.4

Sripada, R. K., Henry, J., Yosef, M., Levine, D. S., Bohnert, K. M., Miller, E., & Zivin, K. (2016). Occupational functioning and employment services use among VA primary care patients with posttraumatic stress disorder. *Psychological Trauma: Theory Research, Practice, and Policy*, *10*(2), 140–143. https://doi.org/10.1037/tra0000241

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, *14*(4), 323–348. https://doi.org/10.1037/a0016973

Taft, C. T., Watkins, L. E., Stafford, J., Street, A. E., & Monson, C. M. (2011). Posttraumatic stress disorder and intimate relationship problems: A meta-analysis. *Journal of Consulting and Clinical Psychology*, *79*(1), 22–33. https://doi.org/10.1037/a0022196

Tolkmitt, F., Helfrich, H., Standke, R., & Scherer, K. R. (1982). Vocal indicators of psychiatric treatment effects in depressives and schizophrenics. *Journal of Communication Disorders*, *15*(3), 209–222. https://doi.org/10.1016/0021-9924(82)90034-X

Trevino, A. C., Quatieri, T. F., & Malyska, N. (2011). Phonologically-based biomarkers for major depressive disorder. *EURASIP Journal on Advances in Signal Processing*, *2011*(1), 42. https://doi.org/10.1186/1687-6180-2011-42

Vanello, N., Guidi, A., Gentili, C., Werner, S., Bertschy, G., Valenza, G., & Scilingo, E. P. (2012). Speech analysis for mood state characterization in bipolar patients. *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, 2104–2107. https://doi.org/10.1109/EMBC.2012.6346375.

Weathers, F. W., Bovin, M. J., Lee, D. J., Sloan, D. M., Schnurr, P. P., Kaloupek, D. G., & Marx, B. P. (2017). The Clinician-Administered PTSD Scale for DSM–5 (CAPS-5): Development and initial psychometric evaluation in military veterans. *Psychological Assessment*, *30*(3), 383–395. https://doi.org/10.1037/pas0000486

Yang, Y., Fairbairn, C., & Cohn, J. F. (2013). Detecting depression severity from vocal prosody. *IEEE Transactions on Affective Computing*, *4*(2), 142–150. https://doi.org/10.1109/T-AFFC.2012.38

Zen, A. L., Whooley, M. A., Zhao, S., & Cohen, B. E. (2012). Post-traumatic stress disorder is associated with poor health behaviors: Findings from the heart and soul study. *Health Psychology*, *31*(2), 194–201. https://doi.org/10.1037/a0025989

Zhuang, X., Rozgić, V., Crystal, M., & Marx, B. P. (2014). Improving speech-based PTSD detection via multi-view learning. *Spoken Language Technology Workshop (SLT), 2014 IEEE*, 260–265. https://doi.org/10.1109/SLT.2014.7078584

Zoladz, P. R., & Diamond, D. M. (2013). Current status on behavioral and biological markers of PTSD: A search for clarity in a conflicting literature. *Neuroscience & Biobehavioral Reviews*, *37*(5), 860–895. https://doi.org/10.1016/j.neubiorev.2013.03.024

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Marmar CR, Brown AD, Qian M, et al. Speech-based markers for posttraumatic stress disorder in US veterans. *Depress Anxiety*. 2019;36:607–616. https://doi.org/10.1002/da.22890