An Emotion-Driven Vocal Biomarker-Based PTSD Screening Tool

Thomas F. Quatieri , *Fellow, IEEE*, Jing Wang , James R. Williamson , Richard DeLaura , Tanya Talkar , Nancy P. Solomon , Stefanie E. Kuchinsky , Megan Eitel, Tracey Brickell, Sara Lippa, Kristin J. Heaton , Douglas S. Brungart, Louis French, Rael Lange, Jeffrey Palmer , *Senior Member, IEEE*, and Hayley Reynolds

Abstract*—*Goal: **This paper introduces an automated post-traumatic stress disorder (PTSD) screening tool that could potentially be used as a self-assessment or inserted into routine medical visits to aid in PTSD diagnosis and treatment. *Methods:* With an emotion estimation algorithm providing arousal (excited to calm) and valence (pleasure to displeasure) levels through discourse, we select regions of the acoustic signal that are most salient for PTSD detection. Our algorithm was tested on a subset of data from the DVBIC-TBICoE TBI Study, which contains PTSD Check List Civilian (PCL-C) assessment scores. *Results:* Speech from low-arousal and positive-valence regions provide the highest discrimination for PTSD. Our model achieved an AUC (area under the curve) of 0.80 in detecting PCL-C ratings, outperforming models with no emotion filtering (AUC = 0.68). *Conclusions:* This result suggests that emotion drives the selection of the most salient temporal regions of an audio recording for PTSD detection.**

***Index Terms*—Emotional digital twin, emotion sensing, neuromotor coordination, PTSD, vocal biomarkers.**

***Impact Statement*—Vocal biomarkers based on temporal regions of low-arousal and positive-valence achieve an area under the curve of 0.80 in detecting PTSD Check List Civilian (PCL-C) ratings.**

Manuscript received 5 December 2022; revised 17 March 2023 and 7 June 2023; accepted 7 June 2023. Date of publication 13 June 2023; date of current version 15 August 2024. This work was supported by the U.S. Army Military Operational Medicine Research Program (MOMRP). The work of Tanya Talkar was supported by NIH-NIDCD Fellowship under Award F31DC019509. The review of this article was arranged by Editor Marianna Laviola. *(Corresponding authors: Thomas Quatieri; Jing Wang.)*

Thomas F. Quatieri and Tanya Talkar are with the MIT Lincoln Laboratory (MIT LL), Lexington, MA 02421 USA, and also with the Speech and Hearing Bioscience and Technology Program, Harvard Medical School, Cambridge, MA 02115 USA (e-mail: quatieri@ll.mit.edu; tanya.talkar@ll.mit.edu).

Jing Wang, James R. Williamson, Richard DeLaura, Jeffrey Palmer, and Hayley Reynolds are with the MIT Lincoln Laboratory (MIT LL), Lexington, MA 02421 USA (e-mail: jing.wang@ll.mit.edu; jrw@ll.mit.edu; richd@ll.mit.edu; jspalmer@ll.mit.edu; hayley@ll.mit.edu).

Nancy P. Solomon, Stefanie E. Kuchinsky, and Douglas S. Brungart are with Walter Reed National Military Medical Center (WRNMMC), Bethesda, MD 20889 USA, and also with the Uniformed Services University of the Health Sciences, Bethesda, MD 20814 USA (e-mail: nancy.p.solomon.civ@health.mil; stefanie.e.kuchinsky.civ@health.mil; douglas.s.brungart.civ@health.mil).

Megan Eitel is with Walter Reed National Military Medical Center (WRNMMC), Bethesda, MD 20889 USA (e-mail: eitelmm@gmaill.com).

Tracey Brickell and Rael Lange are with the Traumatic Brain Injury Center of Excellence, Walter Reed National Military Medical Center (WRNMMC), Bethesda, MD 20814 USA, and with the National Intrepid Center of Excellence, WRNMMC, Bethesda, MD 20814 USA, and with the General Dynamics Information Technology, Silver Spring, MD 20910 USA, and also with the Uniformed Services University of the Health Sciences, Bethesda, MD 20814 USA (e-mail: tbrickell.dvbic@gmail.com; rael.lange@gmail.com).

Sara Lippa is with the National Intrepid Center of Excellence, WRNMMC, Bethesda, MD 20814 USA, and also with the Uniformed Services University of the Health Sciences, Bethesda, MD 20814 USA (e-mail: saralippa@gmail.com).

Kristin J. Heaton is with the U.S. Army Research Institute of Environmental Medicine (USARIEM), Natick, MA 01760 USA (e-mail: kristin.j.heaton.civ@health.mil).

Louis French is with the Traumatic Brain Injury Center of Excellence, Walter Reed National Military Medical Center (WRNMMC), Bethesda, MD 20814 USA, and with the National Intrepid Center of Excellence, WRNMMC, Bethesda, MD 20814 USA, and also with the Uniformed Services University of the Health Sciences, Bethesda, MD 20814 USA (e-mail: louis.m.french.civ@mail.mil).

This article has supplementary downloadable material available at https://doi.org/10.1109/OJEMB.2023.3284798, provided by the authors.

Digital Object Identifier 10.1109/OJEMB.2023.3284798

I. Introduction

POST-traumatic stress disorder (PTSD) is a disorder that can occur after experiencing or witnessing a life-threatening or traumatic event [44]. It is characterized by nightmares or intrusive traumatic memories, avoidance, hypervigilance, anxiety, or depressed mood that persist for more than a month after the trauma and cause significant disruption to daily living and work activities [1]. Although often associated with trauma sustained during military operations, PTSD is prevalent in both civilian and military populations, impacting nearly 12 million adults in the U.S. each year [31]. PTSD can be difficult to diagnose because it is highly individualized and its symptoms can be subtle and transient. In addition, it shares common diagnostic features with other conditions, such as depression, anxiety, and brain injury, and since avoidance is a key feature of PTSD, many patients may lack awareness or understanding of their symptoms or may resist speaking of them with medical providers. Yet many current methods to evaluate PTSD rely on self-report measures and these are vulnerable to symptom exaggeration for some individuals [24].

The ability to objectively and non-intrusively detect PTSD could facilitate diagnostic accuracy and, when used as a self-screening measure, could provide an opportunity for earlier detection and intervention. However, simple, objective, non-invasive tools to screen for PTSD are currently lacking. Such a tool could augment clinical diagnosis and assist in tracking the effectiveness of a therapy. Our vision is a 'digital emotional twin' driving measurement of vocal markers that can be incorporated into a PTSD screening dashboard as a tool for both patients and healthcare providers in civilian and military populations to monitor progress and enable interventions as symptoms change. We can think of the digital emotional twin as, ideally, a personalized computational model of the effect of emotion on speech and language neurophysiology, spanning the human lifecycle, and updated from real-time, individualized data.

A. Benefits of a Vocal-Based Screening Tool

According to current research, psychological screening tools, including self- and clinician-based assessments, can be reasonably effective at identifying PTSD in patients, with self-report assessments such as a PTSD Checklist (PCL) achieving a classification figure of merit score between 0.77–0.80 out of 1.00 [43] using the computed area under the curve (AUC) value from a receiver operating characteristic (ROC) curve [30]. A vocal biomarker-based PTSD screening tool ideally should exceed current self-report and clinical methods in performance. However, the use of vocal biomarkers, despite when somewhat lower performance occurs, provide the benefit of being more easily scalable to a broader population and lend themselves to continuous, nonobtrusive assessment. The other clear benefit is that such a tool, not relying on specialized equipment, aids self-assessment and can easily be adapted to individual differences. For many military and civilian applications, particularly in field-forward environments or less equipped clinics, common modalities such as video recordings and blood assays may not be available or feasible (see **Supplement**).

B. Physiological Underpinnings

With the work of others [28], [39], [41], clinical subjective observations, and our previous work in cognitive state estimation [34], [47] as grounding, we investigate how PTSD may be associated with changes in neuromotor coordination in respiratory, phonatory, and articulatory physiology. Changes in respiratory muscle activity and in other parts of the respiratory tract can affect temporal patterns of inhalation and exhalation for speaking [27], lung pressure for the generation of vocal intensity (perceived as loudness), and the rate of air flow through the larynx affecting voice quality (e.g., breathiness, strain). The respiratory system is highly coordinated with the laryngeal-based subsystem [17], [50], together affecting vibratory characteristics of the vocal folds (e.g., rate for pitch, extent for loudness, and regularity for quality) and certain characteristics of speech articulation (e.g., voiced/voiceless distinction, aspiration for stop consonants). Likewise, respiratory and laryngeal activity are coupled to articulation in the oral and nasal cavities [16], affecting resonance (e.g., oral/nasal consonants, perceived nasality).

Changes in emotion have been shown to have strong vocal correlates. Emotional changes can be quantified along two key dimensions: arousal (i.e., intensity ranging from calm/sleepy to excited/alarmed) and valence (i.e., pleasure to displeasure) [37]. Emotional changes associated with PTSD such as re-experiencing traumatic memories, hypervigilance, or suppression of thoughts or feelings may influence respiratory, phonatory and/or articulatory components of speech production [39]. However, there is varying evidence of the degree to which the subsystems correlate with the arousal and valence elements of emotion (i.e., [7]), or how these emotional components relate to the severity of PTSD over time, and through the varying nature of a discourse.

C. Approach

In deriving predictors of arousal and valence of an individual, i.e., the digital 'emotional twin', we train with separate machine learning models based on a standard emotion dataset. For generalizability to other datasets, we use only time-varying energy in the acoustic signal, a proxy for the respiratory speech subsystem. These continuous estimates of arousal and valence provide the basis of an 'emotion filter' that selects regions of the acoustic signal most salient for PTSD detection. Specifically, motivated by the work of Scherer et al. who found that neutral and positive-polarity questions led to more accurate classification models [39], we select regions of low arousal or positive valence. Additional supporting evidence for this approach comes from the work of Finucane et al. [12] who reported that the most important discriminant to determine if an individual is from a healthy control or is a clinical PTSD case is a lower incidence of positive emotions (happiness) in the later. Patients with depression, which is a closely related condition to PTSD, are also found to show reduced reactivity to positive experiences [2]. Given PTSD subjects have lower incidence of positive emotions, we hypothesize that they are less likely to exhibit speech patterns typically expressed during low arousal and positive valence emotional states. Therefore, by focusing analysis on those parts of speech with lowest arousal and most positive valence within each subject, it is easier to detect PTSD. Given these selected speech regions, we extract speech features and investigate their relative importance across respiratory, phonatory and articulatory speech production subsystems. These features are then used in a machine learning model to predict PTSD severity. A system overview of the predictor is illustrated in Fig. 1.

II. Materials and Methods

A. Datasets

Emotion dataset: The emotion models were trained using the speech portion of the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) dataset [26], which consists of 12 actors and 12 actresses repeating a 2 to 3 seconds-long, neutral-content sentence to convey each of 8 emotion states, namely neutral, calm, happy, sad, angry, fearful, disgusted and surprised. The speakers were instructed to speak once at a normal and then repeat at a strong emotional intensity level. A total of 1440 files (60 trials per actor x 24 actors = 1440) were used for training the models. An acted emotion dataset was selected over a naturalistic emotion dataset because the acted emotion exemplars tend to better reflect prototypical components of an emotional state than those produced spontaneously by non-actors [40]. Thus, acted emotion provides a clearer emotional expression that is more useful for cross-corpus speech emotion recognition [40], [51].

PTSD dataset: To test our detection algorithm, we used a subset of the Defense and Veterans Brain Injury Center/Traumatic Brain Injury Center of Excellence (DVBIC-TBICoE) 15-year Longitudinal TBI Study [25]. In addition to the evaluation of TBI severity and depression, this study measures symptoms of PTSD as part of comorbidity background information. Speech-language assessments took place in a double-walled sound-attenuating booth with the participant seated in front of a video monitor and the examiner visible through the examination window. Recordings were collected onto a laboratory

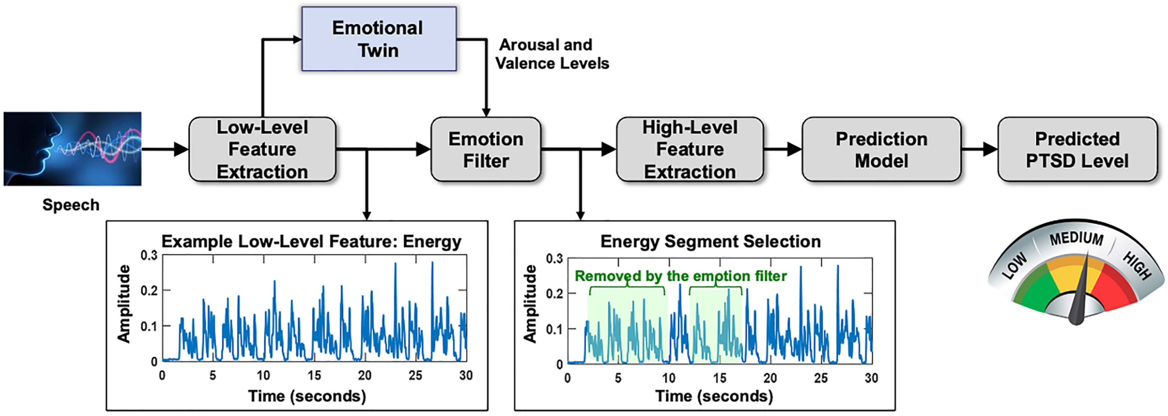

Fig. 1. Schematic of PTSD prediction system overview. Continuous estimates of arousal and valence provide the basis of an 'emotion filter' that selects regions of the acoustic signal most salient for PTSD detection.

computer via a cardioid dynamic microphone (Shure PG48) with a constant mouth-to-microphone distance of 14 cm. Recordings were made through an internal soundcard (RME Hammer-fall DSP Multiface II) using a software audio recorder function (MATLAB 2007). Subjects were asked to perform a series of speech tasks. These tasks included having the subject watch a wordless picture story and retell the story for their free-speech sample. If participants were unable to provide an adequate storytelling sample, they were asked to tell the examiner about their day. Participants were then asked to read a pre-scripted paragraph utilizing a variety of words of increasing length and complexity called "The Caterpillar passage" [52]. This passage was designed to examine segmental and prosodic skills related to speech motor production, and is characterized by contemporary themes with relatively simple reading requirements.

The current analyses included a total of 141 subjects with no history of TBI or with mild TBI who completed the read speech and/or the free speech task. Subjects with Major Depression Disorder (MDD) were not excluded since this condition highly correlates with PTSD. PTSD severity was quantified using the PTSD Checklist-Civilian version (PCL-C) scores [25], [46]. The PCL-C contains three main sections, namely "B", "C" and "D" with 17 questions in total, each of which have response options that range from 1 "Not at all" to 5 "Extremely". The total symptom severity score is the sum of all the responses, thus ranging from 17 to 85. Additional details about this study are found in the **Supplement** and also in Lange et al. [25] with inclusion criteria, group definition, and recruiting procedures.

B. Emotion Estimation

Our emotion models of arousal and valence use univariate statistics of speech-intensity features for simplicity in interpretation and transferability in applying to cross-corpus datasets. Here, speech intensity or energy, an approximate measure of loudness, refers to the time-domain speech-signal envelope. It is estimated using a custom MATLAB script that iteratively filters the speech-signal amplitude until a smooth contour is obtained [22], [36]. This technique captures both the contributions of the respiratory system and resonance-harmonics interactions to amplitude modulation of a speech envelope [22], [42].

The speech-intensity timeseries were first normalized within subject by subtracting the mean and dividing by the standard deviation of all recordings from the same subject. The time derivatives of the speech intensity were then computed using the normalized speech intensity time series. Second, univariate statistics were computed for both normalized speech intensity time series and their time derivatives (using only the voicing regions of the speech segments), including mean, median, mode, standard deviation, range, inter-quartile ranges, skew, kurtosis, minimum and maximum. Lastly, the features were normalized across subjects by subtracting the group mean and dividing by the standard deviation for each feature. In order to reduce feature dimensions, principal component analysis (PCA) was applied to the normalized features, and the optimal number of components were chosen, defined as the point 95% of feature variances are explained.

Using the RAVDESS dataset, two separate machine learning models were trained to detect high emotional arousal and non-negative emotional valence, respectively. The RAVDESS study instructs participants to read sentences in the following emotional states: neutral, calm, happy, sad, angry, fearful, surprise, and disgust. For each emotional state, they are also instructed to express at two levels of emotional intensity (normal, strong). For valence classification, the analysis grouped sad, angry, fearful and disgusted as negative emotions, and the remaining conditions (neutral, calm, happy, surprised) as positive emotions. For arousal classification, the analysis grouped normal intensity as low arousal, and strong intensity as high arousal. Both models used support vector machine (SVM) classification algorithms to compute classification scores (signed distance from the decision boundary for a speech segment to be in a given class). In this analysis, larger classification scores indicate a higher likelihood that the speaker exhibited a stronger emotional intensity or a more positive valence. The scores from the emotional intensity model were used as estimates of arousal state, and the scores from the positive emotion model were used as estimates for valence state. More details of training the predictive emotion models and application of our emotion models to the DVBIC-TBICoE data are given in the **Supplement**.

C. PTSD Detection

As illustrated in Fig. 1, PTSD detection relies on emotion filtering followed by high-level feature extraction and classification. Emotion filtering passes low-level feature segments that correspond to low arousal or positive valence (i.e., a calmer or more positive state of the subject) on to feature processing. Here we describe the feature extraction and classification for PTSD prediction**,** leveraging our previous use of neuromotor coordination-based vocal markers to detect a variety of neurological and psychiatric conditions such as depression, cognitive fatigue, traumatic brain injury, and dementia [34]. The acoustic-derived markers of speech-production subsystems provide a proxy to the underlying muscular coordination that influence the behavioral measures.

Feature extraction: Standard low-level features associated with the various speech subsystems characterize basic properties of the three vocal subsystem components. For the *respiratory* subsystem, the speech envelope is used as a proxy for vocal intensity. For the *phonatory* subsystem, we estimate the fundamental frequency (pitch) and cepstral peak prominence (CPP), which provides a measure of stability of vocal fold vibration. For the *articulatory* subsystem, we used the mel-frequency cepstral coefficients (MFCCs) as a proxy of articulatory movements. The high-level features include representations of correlation structure and dispersion, characterizing levels and complexity of coordination in these subsystems. In applying our emotion filter within each time series, masking was used to exclude speech segments that are associated with either high arousal or negative valence. (Other details of both low- and high-level features are provided in the **Supplement**).

Prediction models: The prediction models described below were created independently in five different cross-validation training folds. Subjects were randomly divided into five testing and training folds containing 20% and 80% of the subjects, respectively. The same number of principal component features were used in all training folds. For the delta-MFCC, delta-envelope, and delta-CPP coordination features, 5 principal components were extracted from each of the eigenvalue feature vectors, which have dimensionality 780, 60, and 60 respectively. The dimensionality is determined by the number of channels (12, 1 and 1) times the number of time delay scales (4, 4 and 4) and the number of delays per scale (15, 15 and 15). For the dispersion features, a single principal component is extracted from the delta-CPP feature vector, which has dimensionality 2 (mean path length and mean distance). The following summarize the emotion-driven ('emotion-filtered') feature selection.

- *Respiration:* 30% of data over time with lowest arousal or highest valence were analyzed for correlation-based delta-envelope features
- *Phonation:* 20% of data over time with lowest arousal or highest valence were analyzed for correlation-based delta-CPP features; 50% of data over time with lowest arousal or highest valence was analyzed for dispersion-based delta-CPP features
- *Articulation:* 20% of data over time with lowest arousal or highest valence was analyzed for correlation-based delta-MFCC features.

Gaussian mixture models (GMMs) and cross-validation training form the basis of classification (See **Supplement** for details). PCL-C prediction results were obtained by training and testing prediction models on read and free speech data from the DVBIC-TBICoE data set.

TABLE I
MODEL PREDICTION PERFORMANCE

Subsystem	Feature	Arousal or valence	Spearman R, AUC (PCLC > 35)		
			Read	Free	Read & Free
Respiration	delta-Env xcorr	arousal	0.18 0.57	0.15, 0.66	0.21, 0.64
	delta-Env xcorr	valence	0.19, 0.70	0.12, 0.61	0.17, 0.66
Phonation	delta-CPP xcorr	arousal	0.26, 0.62	0.25, 0.59	0.32, 0.63
	delta-CPP xcorr	valence	0.30, 0.64	0.21, 0.60	0.32, 0.65
	delta-CPP disp	arousal	0.28, 0.64	0.35, 0.72	0.35, 0.71
	delta-CPP disp	valence	0.34, 0.69	0.36, 0.75	0.39, 0.74
Articulation	delta-MFCC xcorr	arousal	0.35, 0.71	0.27, 0.70	0.38, 0.74
	delta-MFCC xcorr	valence	0.35, 0.73	0.21, 0.66	0.36, 0.74
Fused	-	-	0.38, 0.72	0.40, 0.77	0.47, 0.80

Spearman correlation of prediction scores and actual PCL-C for the eight features sets, on read speech, free speech, and fused across both speech tasks. The feature sets employ different emotion filtering thresholds as described in the text. The best performance of R=0.47 and AUC=0.80 is obtained by fusing across all eight feature sets and both speech tasks.

III. RESULTS

Classification resulted in an area under the curve (AUC) performance of 0.80, based on labeling PCL-C scores > 35 as positive cases. A threshold of 35 was selected to allow for a sufficient number of positive cases ($n = 33$) in doing classification (e.g., increasing to 43.5 and 49.5 result in $n = 17$ and 11, respectively). (A cutoff score near 35 has also been suggested as an appropriate choice for screening a population by a primary care or non-mental health clinician: https://learn.livingwell.org.au/mod/page/view.php?id=65.) As seen in Table I, this result was achieved by fusing across eight feature sets, that consist of four underlying feature sets combined with two methods of emotion filtering, using arousal and valence estimates, and by fusing across predictions from read speech and free speech tasks. We see in Table I the relative importance of speech subsystems (respiration, phonation, articulation) in PCL-C score prediction. We also see that the prediction performance of each subsystem tends to be both feature- and task-dependent. Fusion across all subsystems provides the highest prediction performance in contrast to using any one individual subsystem. Fig. 2 illustrates the relationship between PCL-C scores and fused scores, while Fig. 3 shows the ROC curve (true detections versus false alarms) for detecting PCL-C > 35 with emotion filtering. Various cost-benefit tradeoffs occur as we move along the ROC curve: for example, we achieve 80% detection accuracy at a 25% false alarm rate.

IV. DISCUSSION

The proposed PTSD detection algorithm benefits significantly from the incorporation of emotional valence and arousal filtering (low arousal and positive valence), resulting in an AUC of 0.80. This AUC represents an improvement over no emotional filtering (AUC of 0.68), as well as over an emotional filter using high arousal and negative valence (AUC $= 0.70$). Previous studies of PTSD detection from speech either did not consider emotional state changes [28], [39], [41] or elicited emotion-specific speech via an external stimuli [39]. Our result suggests that a digital emotional twin can be used to select the most salient temporal regions of an audio recording to improve PTSD detection

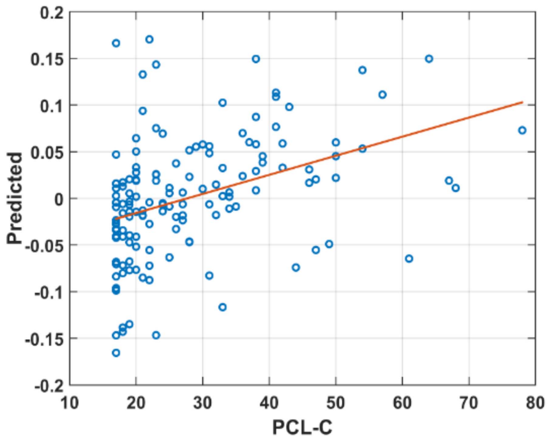

Fig. 2. Prediction/ground truth scatter plot corresponding to the fusion of all eight feature sets. A linear fit is shown as a visual aid while the Spearman correlation yields R = 0.47. The corresponding AUC (Fig. 3) is 0.80.

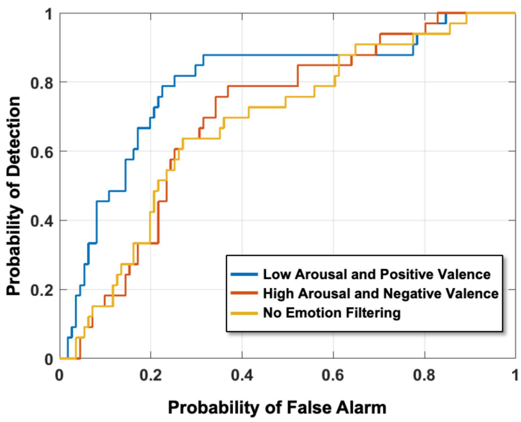

Fig. 3. ROC curves corresponding to the comparison of PTSD detection performance with positive emotion filtering (low arousal & positive valence) (AUC = 0.80), negative emotion filtering (high arousal & negative valence) (AUC = 0.70), and no emotion filtering (AUC = 0.68).

performance. This novel usage of an emotional twin may be extended to improve detections of other psychological conditions.

Our results also provide insight into the relative importance of speech subsystems (respiration, phonation, articulation) for PTSD prediction indicating that this importance is feature- and task-dependent. Furthermore, when using features associated with each subsystem separately, respiration gave the lowest performance while phonation and articulation were overall similar in performance. Nevertheless, the fusion of classifiers that included all three subsystems resulted in the highest performance of AUC equal to 0.80 in detecting the PTSD self-report PCL-C ratings.

Further investigation with additional data is needed to assess if the algorithm can distinguish PTSD from other conditions, such as anxiety disorder, or distinguish subtypes of PTSD. Limitations of the study and a future vision are provided in the **Supplement**.

V. Conclusion

With a level of performance of 0.80 AUC in detecting the PTSD PCL-C ratings, our next step is to validate with broader datasets and incorporate the algorithm into a clinical prototype and evaluate the concept of operations of this non-invasive PTSD detection tool with patients and clinicians. Use cases where the tool is stand-alone or applied synergistertically with standard self-report and clinical assessments will be considered. Algorithm enhancements could include targeted collection of female participants, as well as other under-represented groups in current PTSD datasets, to ensure sensitivity robustness across the entire population. Other future directions include investigating motor coordination across speech subsystems as well as within subsystems, aspects of emotion beyond arousal and valence, the effects of different environmental conditions and platforms, and the refinement of the algorithm to better address specificity concerns with comorbid mental health disorders.

Acknowledgment

The authors would like to thank Dan Hannon, Sophia Yuditskaya and Jonathan Taylor from MIT Lincoln Laboratory for helpful discussions and software guidance.

MIT Lincoln Laboratory (MIT LL), Lexington, MA, USA: Distribution Statement A. Approved for public release. Distribution is unlimited. This material is based upon work supported by the Dept of the Army under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Dept of the Army. Traumatic Brain Injury Center of Excellence, Walter Reed National Military Medical Center (WRNMMC), Bethesda, MD, USA: The views expressed in this manuscript are those of the authors and do not necessarily represent the official policy or position of the Defense Health Agency, Department of Defense, or any other U.S. government agency. This work was prepared under Contract HT0014-22-C-0016 with DHA Contracting Office (CO-NCR) HT0014 and, therefore, is defined as U.S. Government work under Title 17 U.S.C.§101. Per Title 17 U.S.C.§105, copyright protection is not available for any work of the U.S. Government. For more information, please contact dha.TBICOEinfo@health.mil. UNCLASSIFIED. U.S. Army Research Institute of Environmental Medicine (USARIEM), Natick, MA, USA: The opinions or assertions contained herein are the private views of the author(s) and are not to be construed as official or reflecting the views of the Army or the Department of Defense.

References

[1] American Psychiatric Association, *American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM V)*, 4th ed. WA, DC, USA: Amer. Psychiatric Assoc., 2013.

[2] L. M. Bylsma, B. H. Morris, and J. Rottenberg, "A meta-analysis of emotional reactivity in major depressive disorder," *Clin. Psychol. Rev.*, vol. 28, no. 4, pp. 676–691, 2008.

[3] P. Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," *Proc. Inst. Phonetic Sci.*, vol. 17, no. 1193, pp. 97–110, 1993.

[4] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer [computer program]," Version 6.3.01, 1992–2022. Accessed: Jan. 23, 2022. [Online]. Available: http://www.praat.org/

[5] B. J. Borgström and M. S. Brandstein, "Speech enhancement via attention masking network (SEAMNET): An end-to-end system for joint suppression of noise and reverberation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 515–526, 2021.

[6] K. Brady et al., "Multi-modal audio, video and physiological sensor learning for continuous emotion prediction," in *Proc. 6th Int. Workshop Audio/Vis. Emotion Challenge*, 2016, pp. 97–104.

[7] E. F. Briefer, "Vocal communication of emotions," *J. Zool*, vol. 288, pp. 1–20, 2012, doi: 10.1111/j.1469-7998.2012.00920.x.

[8] M. Buller et al., "Gait instability and estimated core temperature predict exertional heat stroke," *Brit. J. Sports Med.*, vol. 56, no. 8, pp. 446–451, 2022.

[9] M. L. Cohen, J. A. Holdnack, P. A. Kisala, and D. S. Tulsky, "A comparison of PHQ-9 and TBI-QOL depression measures among individuals with traumatic brain injury," *Rehabil. Psychol.*, vol. 63, no. 3, pp. 365–371, 2018, doi: 10.1037/rep0000216.

[10] Congressional Research Service, "A guide to U.S. military casualty statistics: Operation freedom's sentinel, operation inherent resolve, operation new dawn, operation Iraqi freedom, and operation enduring freedom," 2015. [Online]. Available: https://apps.dtic.mil/sti/pdfs/AD1169956.pdf

[11] K. R. Dean et al., "Multi-omic biomarker identification and validation for diagnosing warzone-related post-traumatic stress disorder," *Mol. Psychiatry*, vol. 25, pp. 3337–3349, 2020, doi: 10.1038/s41380-019-0496-z.

[12] A. M. Finucane, A. Dima, N. Ferreira, and M. Halvorsen, "Basic emotion profiles in healthy, chronic pain, depressed and PTSD individuals," *Clin. Psychol. Psychother.*, vol. 19, no. 1, pp. 14–24, 2012.

[13] Foundations: Trauma Checklist (PCL-C), "Living well," 2018. Accessed: Nov. 14, 2022. [Online]. Available: https://learn.livingwell.org.au/mod/page/view.php?id=65#:∼:text=The%20PCL%2DC%20a

[14] R. Fraile and J. I. Godino-Llorente, "Cepstral peak prominence: A comprehensive analysis," *Biomed. Signal Process. Control*, vol. 14, pp. 42–54, 2014.

[15] J. M. Goldstein, M. Jerram, B. Abbs, S. Whitfield-Gabrieli, and N. Makris, "Sex differences in stress response circuitry activation dependent on female hormonal cycle," *J. Neurosci.*, vol. 30, no. 2, pp. 431–438, 2010.

[16] V. L. Gracco and A. Lofqvist, "Speech motor coordination and control: Evidence from lip, jaw, and laryngeal movements," *J. Neurosci.*, vol. 14, no. 11, pp. 6585–6597, 1994.

[17] P. Gramming, J. Sundberg, S. Ternström, R. Leanderson, and W. H. Perkins, "Relationship between changes in voice pitch and loudness," *J. Voice*, vol. 2, no. 2, pp. 118–126, 1988.

[18] Y. D. Heman-Ackah, D. D. Michael, and G. Jr, and S. G., "The relationship between cepstral peak prominence and selected parameters of dysphonia," *J. Voice*, vol. 16, no. 1, pp. 20–27, 2002.

[19] J. Hillenbrand, R. A. Cleveland, and R. L. Erickson, "Acoustic correlates of breathy vocal quality," *J. Speech, Lang., Hear. Res.*, vol. 37, no. 4, pp. 769–778, 1994.

[20] C. W. Hoge, L. A. Riviere, J. E. Wilk, R. K. Herrell, and F. W. Weathers, "The prevalence of post-traumatic stress disorder (PTSD) in US combat soldiers: A head-to- head comparison of DSM-5 versus DSM-IV-TR symptom criteria with the PTSD Checklist," *Lancet Psychiatry*, vol. 1, pp. 269–277, 2014, doi: 10.1016/s2215-0366(14)70235-4.

[21] P. Homan et al., "Neural computations of threat in the aftermath of combat trauma," *Nature Neurosci.*, vol. 22, no. 3, pp. 470–476, 2019.

[22] R. L. Horwitz-Martin, T. F. Quatieri, E. Godoy, and J. R. Williamson, "A vocal modulation model with application to predicting depression severity," in *Proc. IEEE 13th Int. Conf. Wearable Implantable Body Sensor Netw.*, 2016, pp. 247–253.

[23] J. L. Judkins, B. A. Moore, T. L. Collette, W. J. Hale, A. L. Peterson, and S. B. Morissette, "Incidence rates of posttraumatic stress disorder over a 17-year period in active duty military service members," *J. Traumatic Stress*, vol. 33, no. 6, pp. 994–1006, 2020, doi: 10.1002/jts.22558.

[24] R. T. Lange, T. A. Brickell, S. M. Lippa, and L. M. French, "Clinical utility of the neurobehavioral symptom inventory validity scales to screen for symptom exaggeration following traumatic brain injury," *J. Clin. Exp. Neuropsychol.*, vol. 37, no. 8, pp. 853–862, 2015.

[25] R. T. Lange et al., "Long-term neurobehavioural symptom reporting following mild, moderate, severe, and penetrating traumatic brain injury in U.S. military service members," *Neuropsychol. Rehabil.*, vol. 30, no. 9, pp. 1762–1785, 2020, doi: 10.1080/09602011.2019.1604385.

[26] S. R. Livingstone and F. A. Russo, "The Ryerson audio-visual database of emotional speech and song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English," *PLoS One*, vol. 13, no. 5, 2018, Art. no. e0196391, doi: 10.1371/journal.pone.0196391.

[27] A. M. MacLarnon and G. P. Hewitt, "The evolution of human speech: The role of enhanced breathing control," *Amer. J. Phys. Anthropol.*, vol. 109, pp. 341–363, 1999, doi: 10.1002/(SICI)1096-8644(199907)109:3<341::AID-AJPA5>3.0.CO;2-2.

[28] C. R. Marmar et al., "Speech-based markers for posttraumatic stress disorder in US veterans," *Depression Anxiety*, vol. 36, no. 7, pp. 607–616, 2019, doi: 10.1002/da.22890.

[29] R. A. Morey et al., "Amygdala volume changes in posttraumatic stress disorder in a large case-controlled veterans group," *Arch. Gen. Psychiatry*, vol. 69, no. 11, pp. 1169–1178, 2012.

[30] F. S. Nahm, "Receiver operating characteristic curve: Overview and practical use for clinicians," *Korean J. Anesthesiol.*, vol. 75, no. 1, pp. 25–36, 2022, doi: 10.4097/kja.21209.

[31] National Comorbidity Survey, "NCS-R appendix tables: Table I. Lifetime prevalence of DSM-IV/WMH-CIDI disorders by sex and cohort. Table II. Twelve-month prevalence of DSM-IV/WMH-CIDI disorders by sex and cohort," 2005. [Online]. Available: www.hcp.med.harvard.edu/ncs/publications.php

[32] R. H. Pietrzak et al., "Amygdala-hippocampal volume and the phenotypic heterogeneity of posttraumatic stress disorder: A cross-sectional study," *Jama Psychiatry*, vol. 72, no. 4, pp. 396–398, 2015.

[33] T. F. Quatieri, *Discrete-time Speech Signal Processing: Principles and Practice*. Noida, India: Pearson Educ. India, 2002.

[34] T. F. Quatieri et al., "Multimodal biomarkers to discriminate cognitive state," *Role Technol. Clin. Neuropsychol.*, vol. 9, 2017, Art. no. 409.

[35] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digit. Signal Process.*, vol. 10, no. 1-3, pp. 19–41, 2000.

[36] A. Roebel and X. Rodet, "Efficient spectral envelope estimation and its application to pitch shifting and envelope preservation," in *Proc. Int. Conf. Digit. Audio Effects*, 2005, pp. 30–35.

[37] J. A. Russell, "A circumplex model of affect," *J. Pers. Social Psychol.*, vol. 39, no. 6, 1980, Art. no. 1161.

[38] K. R. Scherer, "Vocal communication of emotion: A review of research paradigms," *Speech Commun.*, vol. 40, pp. 227–256, 2003.

[39] S. Scherer, G. Stratou, J. Gratch, and L. Morency, "Investigating voice quality as a speaker-independent indicator of depression and PTSD," in *Proc. Interspeech*, 2013, pp. 1–5.

[40] B. Schuller, Z. Zhang, F. Weninger, and G. Rigoll, "Selecting training data for cross-corpus speech emotion recognition: Prototypicality vs. generalization," in *Proc. Afeka-AVIOS Speech Process. Conf.*, 2011, pp. 1–4.

[41] K. Schultebraucks, V. Yadav, A. Y. Shalev, G. A. Bonanno, and I. R. Galatzer-Levy, "Deep learning-based classification of posttraumatic stress disorder and depression following trauma utilizing visual and auditory markers of arousal and mood," *Psychol. Med.*, vol. 52, no. 5, pp. 957–967, 2022, doi: 10.1017/S0033291720002718.

[42] T. Talkar et al., "Assessment of speech and fine motor coordination in children with autism spectrum disorder," *IEEE Access*, vol. 8, pp. 127535–127545, 2020, doi: 10.1109/ACCESS.2020.3007348.

[43] Q. Q. Tiet, K. K. Schutte, and Y. E. Leyva, "Diagnostic accuracy of brief PTSD screening instruments in military veterans," *J. Substance Abuse Treat.*, vol. 45, no. 1, pp. 134–142, 2013, doi: 10.1016/j.jsat.2013.01.010.

[44] B. V. der Kolk, "Posttraumatic stress disorder and the nature of trauma," *Dialogues Clin. Neurosci.*, vol. 21, 2022, Art. no. 22034447, doi: 10.31887/DCNS.2000.2.1/bvdkolk.

[45] J. Wang, H. Zhao, and M. J. Girgenti, "Posttraumatic stress disorder brain transcriptomics: Convergent genomic signatures across biological sex," *Biol. Psychiatry*, vol. 91, no. 1, pp. 6–13, 2022, doi: 10.1016/j.biopsych.2021.02.012.

[46] F. W. Weathers, B. T. Litz, D. S. Herman, J. A. Huska, and T. M. Keane, "The PTSD checklist (PCL): Reliability, validity, and diagnostic utility," in *Proc. Annu. Conv. Int. Soc. Traumatic Stress Stud.*, 1993, pp. 1–9.

[47] J. R. Williamson, D. Young, A. A. Nierenberg, J. Niemi, B. S. Helfer, and T. F. Quatieri, "Tracking depression severity from audio and video based on speech articulatory coordination," *Comput. Speech Lang.*, vol. 55, pp. 40–56, 2019.

[48] J. H. Wortmann et al., "Psychometric analysis of the PTSD checklist-5 (PCL-5) among treatment-seeking military service members," *Psychol. Assessment*, vol. 28, no. 11, pp. 1392–1403, 2016, doi: 10.1037/pas0000260.

[49] R. Xu et al., "A voice-based automated system for PTSD screening and monitoring," *Stud. Health Technol. Inform.*, vol. 173, pp. 552–558, 2012.

[50] Z. Zhang, "Respiratory laryngeal coordination in airflow conservation and reduction of respiratory effort of phonation," *J. Voice*, vol. 30, no. 6, 2016, Art. no. 760-e7.

[51] J. Gideon, M. G. McInnis, and E. M. Provost, "Improving cross-corpus speech emotion recognition with adversarial discriminative domain generalization (ADDoG)," *IEEE Trans. Affect. Comput.*, vol. 12, no. 4, pp. 1055–1068, Oct.–Dec. 2021, doi: 10.1109/TAFFC.2019.2916092.

[52] R. Patel et al., "'The caterpillar': A novel reading passage for assessment of motor speech disorders," *Amer. J. Speech-Lang. Pathol.*, vol. 22, no. 1, pp. 1–9, 2013, doi: 10.1044/1058-0360(2012/11-0134).