# Evaluation of Voice Acoustics as Predictors of Clinical Depression Scores

***Nik Wahidah Hashim, *Mitch Wilkes, †Ronald Salomon, †Jared Meggs, and ‡Daniel J. France,** *†‡*Nashville, Tennessee*

**Summary: Objective.** The aim of the present study was to determine if acoustic measures of voice, characterizing specific spectral and timing properties, predict clinical ratings of depression severity measured in a sample of patients using the Hamilton Depression Rating Scale (HAMD) and Beck Depression Inventory (BDI-II).

**Study Design.** This is a prospective study.

**Methods.** Voice samples and clinical depression scores were collected prospectively from consenting adult patients who were referred to psychiatry from the adult emergency department or primary care clinics. The patients were audio-recorded as they read a standardized passage in a nearly closed-room environment. Mean Absolute Error (MAE) between actual and predicted depression scores was used as the primary outcome measure.

**Results.** The average MAE between predicted and actual HAMD scores was approximately two scores for both men and women, and the MAE for the BDI-II scores was approximately one score for men and eight scores for women. Timing features were predictive of HAMD scores in female patients while a combination of timing features and spectral features was predictive of scores in male patients. Timing features were predictive of BDI-II scores in male patients.

**Conclusion.** Voice acoustic features extracted from read speech demonstrated variable effectiveness in predicting clinical depression scores in men and women. Voice features were highly predictive of HAMD scores in men and women, and BDI-II scores in men, respectively. The methodology is feasible for diagnostic applications in diverse clinical settings as it can be implemented during a standard clinical interview in a normal closed room and without strict control on the recording environment.

**Key Words:** depression–severity–prediction–voice–acoustics.

## INTRODUCTION

Psychomotor retardation and cognitive disturbances are characteristic features of major depressive disorder.[1,2] These symptoms occur early in the course of the disorder and may present as disturbances in speech, facial expression, fine motor behavior, and gross locomotor activity, and as deficits in cognitive functioning such as memory, attention, executive functioning, and psychomotor speed. Standard clinical rating scales of depression typically include no more than one or two items to assess psychomotor symptoms. For example, only 2 items in the 17-item Hamilton Depression Rating Scale (HAMD) address these symptoms. The Salpetriere Retardation Rating Scale,[3] the CORE Index,[4] and the Motor Agitation and Retardation Scale[5] have been designed specifically to evaluate psychomotor retardation in depression, but these scales are not widely used in emergency medicine.

The exponential growth of computing power paired with the decreasing costs of consumer electronics is creating new opportunities to investigate voice acoustics for their diagnostic utility.[6–9] The emergence of biometric devices, including but not limited to acoustic analyzers, are well suited to measure the magnitude of psychomotor retardation in depressed depression. Objective measures of voice (eg, fundamental frequency and speech pause time), gross motor activity (eg, wrist flexion and walking), fine motor activity (eg, drawing tasks, eye movements, and facial movements), and cognitive measures (eg, slow ideation and reaction time) have been studied for decades and have shown varying degrees of promise as diagnostic measures for depression.[5,10] However, due to constraints associated with methodology (ie, special testing equipment often required), logistics (ie, space to conduct studies often not available in clinical settings), and analysis (ie, long delays between data and reporting of results), the analysis and application of objective measures have been predominantly confined to laboratory studies.

Of these candidate objective measures, changes in the paralinguistic properties of speech have garnered sustained investigation by researchers as potential biomarkers of depression and depression severity.[11] Paralinguistic refers to the acoustic properties of the voice signal rather than the content or meaning of the speech. Among the distinctive speech patterns associated with depression are decreases in intonation, stress, loudness, inflection, intensity and speech rate, sluggishness in articulation, monotony, and lack in vitality.[12,13] These characteristics correlate with the changes occurring in the speech production mechanism by affecting the respiratory, laryngeal, resonance, and articulatory subsystems that are encoded in the acoustic signal. Acoustic studies of "depressed speech" have included analyses

Accepted for publication June 16, 2016.

Conflicts of interest: The contributing authors have no potential conflicts of interest.

Author contribution statement: NWH, MW, and RS conceived the study and wrote the study protocol and institutional review board application. MW, RS, and DJF mentored NWH during the project, its analysis, and in the dissemination of results. MW and RF supervised the conduct of the pilot study and data collection. RS and JM undertook the recruitment and consenting of patients. NWH managed the data, including quality control. MW and RS provided statistical advice on study design, and NWH analyzed the data. NWH and DJF drafted the manuscript, and all authors contributed substantially to its revision. NWH takes responsibility for the paper as a while.

From the *Department of Electrical Engineering, Vanderbilt University, Nashville, Tennessee; †Department of Psychiatry, Vanderbilt University School of Medicine, Nashville, Tennessee; and the ‡Department of Anesthesiology, Vanderbilt University School of Medicine, Nashville, Tennessee.

Address correspondence and reprint requests to Daniel J. France, Department of Anesthesiology, Vanderbilt University School of Medicine, Nashville, TN 37212. E-mail: Dan.france@vanderbilt.edu

Journal of Voice, Vol. 31, No. 2, pp. 256.e1–256.e6

0892-1997



http://dx.doi.org/10.1016/j.jvoice.2016.06.006

of prosody (eg, pitch, energy, and speaking rate),[14,15] spectral features (eg, power spectral density [PSD], formants and their associated bandwidth,[15–17] glottal features,[12,18] and Mel-Frequency Cepstral Coefficient [MFCC]).[19]

Several studies have demonstrated that specific unidimensional acoustical measures of voice strongly correlate with depression scores derived from clinical rating scales.[20–24] Recent studies have produced preliminary evidence that acoustic measures are effective in discriminating between depressive states, including dysthymia, major depression, and high-risk suicidal states.[15,17,19,25]

The need to improve the detection of objective markers of depression, disease severity, and suicidal risk is paramount. According to the National Institute of Mental Health, the 12-month prevalence of major depressive disorder is 6.9%, affecting 16 million U.S. adults.[26] Of these cases, 30% (2% of the U.S. adult population) are classified as severe. A recent study found that only 2% of patients presenting with nonpsychiatric-related chief complaints who were positively prescreened for mental illness were subsequently identified by emergency medicine physicians in a Level I inner-city trauma center and emergency department (ED).[27]

The inability to reliably diagnose depression in ED is especially problematic for[23] suicide risk prevention because the risk of suicide is 20 times higher in individuals diagnosed with depression than it is in the general population.[28] The fact that suicide victims frequently visit health services in the weeks before death makes the need for improved detection of depression and suicidal risk even more imperative.[29] Validated clinical rating scales, such as the HAMD and Beck Depression Inventory (BDI-II), have improved the screening of depression and suicide ideation by nonpsychiatric clinicians, but these scales have not been shown to effectively predict suicide in individual patients.[30]

## METHODS

The study protocol was submitted to and approved by the Vanderbilt University's Institutional Review Board. Voice samples and HAMD or BDI-II scores were collected prospectively from adult patients who were referred to the study's PI (Salomon) from Vanderbilt's ED and primary care clinics and provided written informed consent to participate in the study.

### Participants

The sample (see Table 1) included adult patients diagnosed as high-risk depressed, nonsuicidal depressed, or remitted after presenting to the Vanderbilt ED. Patients under the influence of alcohol, toxicity, or experiencing respiratory problems such as shortness of breath were excluded. Participants were aged 22–62 years old (43.3 ± 10.2); male patients were younger (39.9 ± 10.3) than female patients (45.3 ± 10.1; *t* test, two-tailed, $P = 0.05$). Male patients had statistically significant higher HAMD scores ($P = 0.04$) but did not differ from female patients on the basis of BDI-II scores (Table 2).

**TABLE 1.**
**Summary of Participants**

| Gender | N | Age | HAMD | BDI-II |
|---|---|---|---|---|
| Male | 44 | 39.9 ± 10.3 | 17.2 ± 5.0 | 25.5 ± 15.8 |
| Female | 72 | 45.3 ± 10.1 | 10.2 ± 3.8 | 26.8 ± 4.0 |

**TABLE 2.**
**Depression Scale Administration by Gender**

| Depression Scale | Male Reading | Female Reading |
|---|---|---|
| Hamilton Depression Rating Scale | 9 | 14 |
| Beck Depression Inventory | 35 | 58 |

### Acquisition of voice samples

The patients were audio-recorded as they were interviewed by nonresearch clinicians and then while they read the standardized "rainbow passage" in a nearly closed-room environment to minimize the disturbance of background noise. The "rainbow passage" was selected because it contains every sound in the English language and is considered to be phonetically balanced with the ratios of assorted phonemes similar to the ones in normal speech.[31]

Audio acquisitions were made using a high-quality cardioid microphone with a frequency response of 40 Hz to 20 kHz, a laptop computer, and a digital audio interface. *PROTools* (Avid, Burlington, MA) *LE* recording software was used for digital audio editing. All voice samples were digitized using a 32-bit analog to digital converter at 44.1 kHz sampling rate. Microphone distance from the patient was 4–6 feet. Recording sensitivity was adjusted while participants counted aloud to 20. Before analysis, the recordings were edited using a free audio digital editor to remove any identifying information and undesirable sounds such as voices other than the patient's, sneezing, coughing, and door slams.

### Depression severity and suicidal risk assessment

The HAMD is the most common diagnostic tool used to measure the severity of depression and suicidal risk in inpatient populations. The HAMD assessment has also been considered as the primary standard for determining suicidal risk. It contains 17-item questionnaires including one item on suicidal thoughts and uses rating scales that can be evaluated only by trained clinicians. Clinicians rely on their intuitions during evaluation and determining the ratings for the provided questionnaires. The generally accepted opinions by clinicians on the interpretation of the total HAMD scores are that scores 0–7 show no presence of depression, 8–13 indicate mild depression, 14–18 indicate moderate depression, 19–22 indicate severe depression, and scores over 22 indicate very severe depression.[32] For a single suicide item, patients scoring 2 or higher were found to be 4.9 times more likely to die by suicide.[33] Even though it has been found to be reliable, the application of this clinician-administered instrument is time-consuming and requires extensive effort by clinicians to obtain repeated comprehensive evaluations. There is a risk of improper assessment due to an accidental failure to inquire about specific information relating to suicide risk and the clinician's lack of well-defined conceptual clarity