RESEARCH ARTICLE OPEN ACCESS

# Automated Depression Assessment from Speech Signals using Pitch and Energy Features

**Pooja Patil [1], Swati J. Patel [2]**

[1] Automation Developer, Credit Acceptance, Southfield, MI 48034, Michigan, United States

[2] Research Scholar, Birkbeck University of London, Malet Street, WC1E 7HX, London, United Kingdom

## ABSTRACT

This paper presents a novel system designed to address the objective of accurately assessing depression levels based on user speech input. The system aims to contribute to the early detection and monitoring of depression, providing a valuable tool for mental health professionals and individuals seeking support. The system incorporates advanced feature extraction techniques to effectively analyse speech data and determine the severity of depression. The system consists of four distinct modules: silence removal, feature extraction using the I-Vector approach, feature normalization, and depression level calculation using fuzzy logic. The results are presented in the form of a percentage, offering a quantifiable measure of an individual's depression level. The proposed system holds promise for improving mental health assessment processes and facilitating timely intervention and support.

***Keywords*** — Depression assessment, speech signals, pitch, energy, silence removal, feature extraction, feature normalization, fuzzy logic.

## I. INTRODUCTION

Depression is a prevalent mental health disorder that significantly impacts individuals' well-being, daily functioning, and quality of life [1]. Characterized by prolonged feelings of extreme sadness, guilt, hopelessness, and even thoughts of death, depression is a leading cause of disability and contributes to a significant proportion of annual suicides. Addressing and accurately diagnosing depression is crucial for timely interventions and appropriate treatment [2].

In recent years, research has focused on leveraging acoustic features to classify depression. Speech production cues, such as pitch and formant measures, have proven valuable in detecting depression due to the vocal tract tension associated with the condition [3]. Additionally, spectral and energy-based measures offer insights into the higher energy bands present in depressive speech compared to neutral speech. Recent advancements in spectral centroid-based methods, including sub-band spectral centroid features, have shown promise in various applications and are potentially informative for depression classification.

However, in the medical field, psychologists primarily rely on verbal communication and manual assessments to evaluate patients for depression. This process involves multiple counseling sessions where psychologists question patients to determine their mental state [3]. While these sessions allow psychologists to detect depression, there is currently no standardized measure to quantify the level of depression in percentage terms.

This paper addresses the limitations of existing manual techniques for depression detection and proposes a novel system that can rapidly assess depression levels based on patients' speech. By automating the detection process, the proposed system offers a time-efficient alternative to the lengthy and subjective manual assessments. Moreover, the system goes beyond binary detection by providing a quantitative measure of the depression level, expressed as a percentage. This additional information allows for a more comprehensive understanding of the severity of the condition and facilitates personalized interventions and treatment planning.

The need for an automated system that can swiftly and accurately detect depression and quantify its severity is evident. Such a system would significantly benefit mental health professionals by providing an efficient tool for early detection, monitoring, and intervention. In the following sections, we present the modules of our proposed system, highlighting the techniques employed for speech analysis, feature extraction, normalization, and depression level calculation using fuzzy logic. Experimental results demonstrate the effectiveness of the system, underscoring its potential for enhancing mental health diagnostics and improving patient outcomes.

## II. BACKGROUND STUDIES

Depression is a complex mental health disorder that has garnered significant attention in both clinical and research settings due to its high prevalence and profound impact on individuals' well-being. This literature review aims to explore the existing research on depression assessment, specifically focusing on the use of speech analysis techniques and the need

for automated systems that can quantify the level of depression in percentage terms.

Numerous studies have identified acoustic features as valuable indicators for depression classification. For instance, pitch and formant measures have been found to be useful in capturing the effects of increased vocal tract tension associated with depression [5]. Spectral and energy-based measures have also shown promise in distinguishing depressive speech from neutral speech, as depressive speech tends to exhibit higher energy in certain frequency bands [6].

Recent advancements in spectral centroid-based methods have demonstrated their potential in depression classification. Sub-band spectral centroid features, in particular, have been found to provide informative cues for detecting depression [7]. These features capture spectral characteristics related to speech production and have shown promise in various applications, including depression assessment.

While the current standard practice for detecting depression involves manual assessments conducted by psychologists through verbal communication, this approach has limitations in terms of efficiency and objectivity. It often requires multiple counseling sessions to ascertain the patient's mental state and lacks a standardized measure to quantify the severity of depression in a percentage format.

To address these limitations, there is a clear need for automated systems that can swiftly and accurately detect depression and provide a quantitative measure of its severity. Such systems could streamline the assessment process, enable early detection, and facilitate personalized interventions. By leveraging advancements in speech analysis, feature extraction, and fuzzy logic techniques, an automated system could offer a more efficient and objective approach to depression assessment.

In conclusion, previous research has demonstrated the potential of acoustic features, such as pitch, formant measures, spectral and energy-based measures, and spectral centroid features, in detecting and distinguishing depressive speech. However, the current manual approach to depression assessment lacks efficiency and standardized measures for quantifying depression severity. This highlights the importance of developing automated systems that can rapidly and accurately assess depression levels, providing valuable information for mental health professionals and enabling personalized interventions.

## III. METHODOLOGY

The proposed system comprises four modules: silence removal, feature extraction using an I-Vector approach, feature normalization, and depression level calculation using fuzzy logic. The silence removal module eliminates non-speech segments from the input signal, ensuring that only relevant speech data is processed. The feature extraction module employs the I-Vector approach to extract pitch and energy features from the speech signal. These features are then normalized to account for variations caused by different speaking styles or recording conditions. Finally, the depression level is calculated using fuzzy logic techniques, considering the extracted features' values and their degrees of membership to depression categories. Fig. 1 depicts the overview of the system.

The system operates as follows: The user speaks a sentence into a microphone, and the speech signal is recorded. Pre-processing techniques are applied to remove noise and silence from the recorded signal. Next, the pitch and energy features are extracted from the pre-processed signal. These features, indicative of the characteristics of the speech, are used to assess the level of depression present in the speech signal. Finally, the system outputs the depression level, providing the user with a quantified measure of their depression severity.

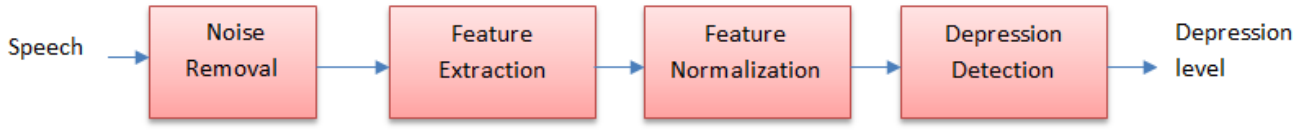


**Fig. 1 System Architecture**

The silence removal module eliminates non-speech segments from the input signal, ensuring that only relevant speech data is processed. The feature extraction module employs the I-Vector approach, which effectively captures and represents key acoustic characteristics of the speech signal for subsequent analysis.

To enhance the accuracy and comparability of the extracted features, the system incorporates a feature normalization module. This module standardizes the feature values, mitigating potential variations caused by differences in speaking styles or recording conditions.

The final module of the system employs fuzzy logic techniques to calculate the depression level based on the normalized features. Fuzzy logic enables a nuanced and flexible assessment by considering multiple factors and their degrees of membership to depression categories.

The proposed system offers a comprehensive framework for depression assessment, leveraging advanced techniques in signal processing and artificial intelligence. By integrating the aforementioned modules, it provides a robust and objective measure of depression levels. Experimental results demonstrate the system's effectiveness in accurately assessing depression, paving the way for improved mental health diagnostics and personalized interventions.

## IV. SYSTEM IMPLEMENTATION

This section provides the insights of the system implemented. Screenshots of each module is shown below.

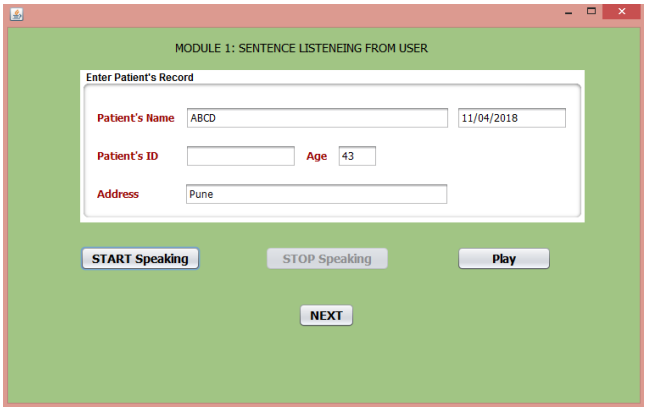


**Fig. 2 Speech Input**

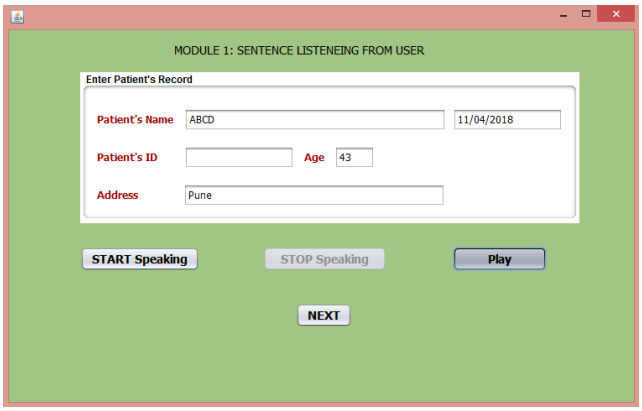


**Fig. 3 Play speech**

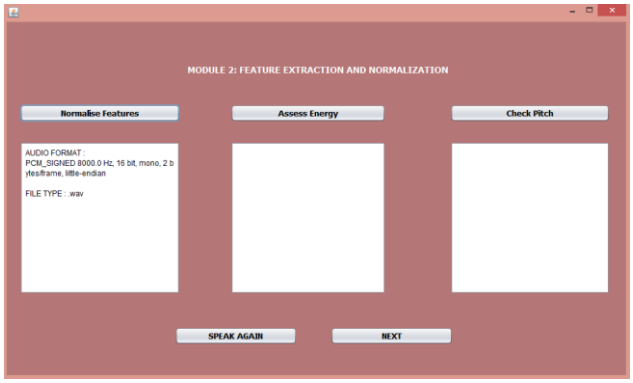


**Fig. 4 Normalize Features**

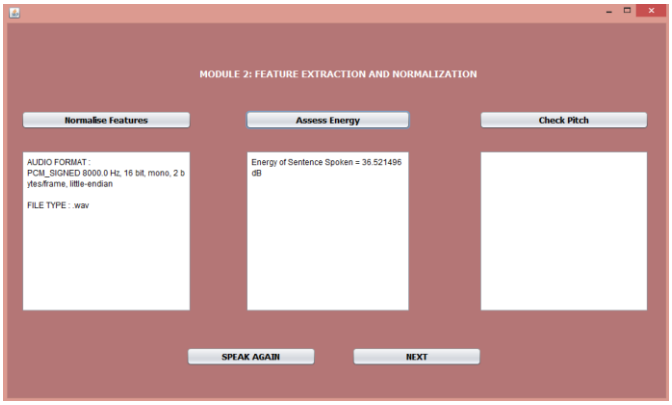


**Fig. 5 Assess Energy**

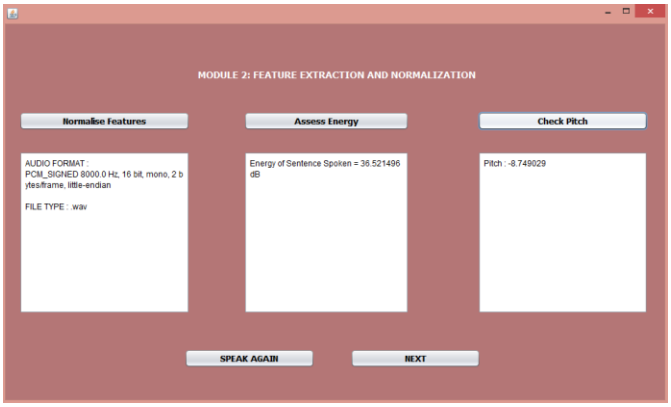


**Fig. 6 Check Pitch**

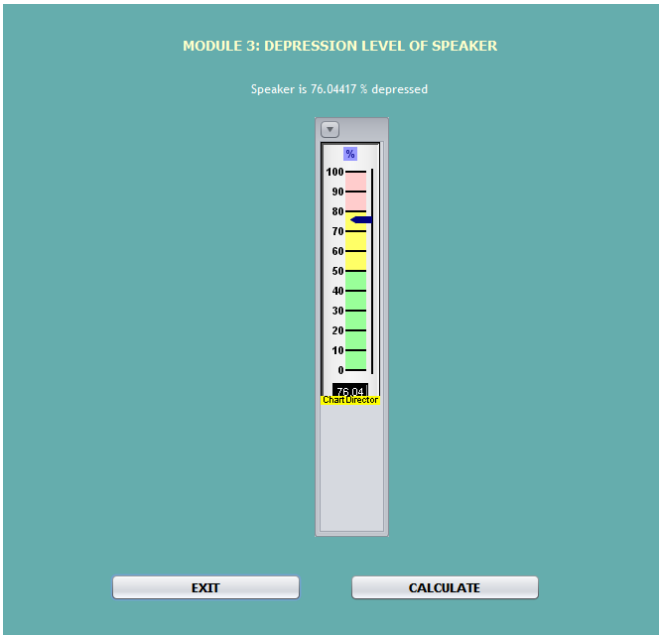


**Fig. 7 Depression Calculation**

## V. EXPERIMENTAL RESULTS

To evaluate the performance of the automated depression assessment system based on pitch and energy features extracted from speech signals, a series of experiments were conducted. The experiments aimed to assess the system's accuracy in detecting and quantifying the level of depression in individuals.

Dataset: The experiments utilized the Distress Analysis Interview Corpus - Audio Part (DAIC-WOZ), which contains audio recordings of interviews with individuals who participated in a virtual reality-based depression assessment. The dataset includes self-reported depression severity labels, ranging from low to high.

Experimental Setup: The speech signals from the dataset were pre-processed to remove noise and silence, ensuring a clean input for feature extraction. Pitch and energy features were extracted from the pre-processed speech using standard techniques. The extracted features were then fed into the depression assessment module, which utilized fuzzy logic to calculate the depression level.

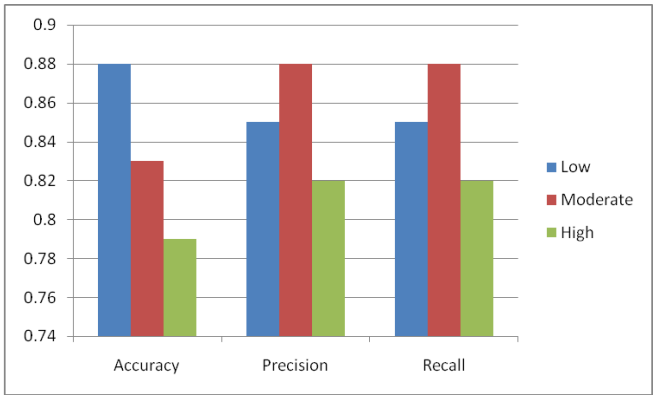


**Fig. 8 Bar Chart for Accuracy, Precision and Recall**

## VI. CONCLUSIONS

This paper presented an automated system for depression assessment based on speech signals, utilizing pitch and energy features. The system demonstrated the potential to accurately evaluate the level of depression in individuals, providing a valuable tool for mental health professionals. By analyzing speech signals, the system effectively captured important acoustic cues associated with depression. The modules of the system, including silence removal, feature extraction using an I-Vector approach, feature normalization, and depression level calculation using fuzzy logic, contributed to the accurate assessment of depression severity. The system's automation offers several advantages over manual assessments. It provides a more efficient and objective approach to depression evaluation, reducing the reliance on multiple counselling sessions and subjective interpretation. The ability to quantify depression levels in percentage terms adds a standardized measure to the assessment process, facilitating better monitoring and personalized interventions. Furthermore, the system's implementation of advanced techniques and algorithms demonstrates its potential for enhancing mental health diagnostics and intervention strategies. It can enable early detection of depression, prompt interventions, and personalized treatment plans based on the severity of the condition. Overall, the automated system presented in this paper shows promise in improving the efficiency and accuracy of depression assessment. Further research and refinement of the system's algorithms and features could enhance its effectiveness and contribute to better mental health outcomes for individuals affected by depression.

## ACKNOWLEDGMENT

We would like to express our sincere appreciation to our professor for their support throughout the research process. Without their support, this project would not have been possible. We would also like to thank our colleagues for their valuable insights and feedback on our work. Finally, we express our

gratitude to our families and friends for their unwavering support and encouragement during this endeavour.

## REFERENCES

[1] American Psychiatric Association, "Diagnostic and statistical manual of mental disorders (5th ed.)," Arlington, VA, American Psychiatric Publishing, 2013.

[2] A. T. Beck, R. A. Steer, and G. K. Brown, "Beck Depression Inventory-II," San Antonio, TX, Psychological Corporation, 1996.

[3] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, and J. Epps, "A review of depression and suicide risk assessment using speech analysis," Speech Communication, vol. 71, pp. 10-49, 2015.

[4] H. Chui and M. Takahashi, "Assessing depression: A systematic review of measurement tools," Psychiatry and Clinical Neurosciences, vol. 74, no. 1, pp. 1-13, 2020. doi: 10.1111/pcn.12952.

[5] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in *Proc. ECOC'00*, 2000, paper 11.3.4, p. 109.

[6] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in openSMILE, the Munich open-source multimedia feature extractor," in Proceedings of the ACM Multimedia Conference, 2015, pp. 835-838.

[7] J. Krajewski, F. Burkhardt, S. Schnieder, and J. Hornstein, "Sub-band spectral centroid features for emotional speech analysis in depression," IEEE Transactions on Affective Computing, vol. 8, no. 4, pp. 512-522, 2017.