Exploring the Complexity of Parkinson's Patient Speech for Depression Detection task: A Qualitative Analysis

Barbara Ruvolo*, Tilak Purohit*†, Bogdan Vlasenko*, Juan Rafael Orozco-Arroyave‡, Mathew Magimai.-Doss*
*Idiap Research Institute, Martigny, Switzerland
†EPFL, École polytechnique fédérale de Lausanne, Switzerland
‡GITA Lab, Universidad de Antioquia, Colombia
{tilak.purohit, mathew.magimaidoss}@idiap.ch

***Abstract*—This study examines the acoustic features of Parkinson's Disease (PD) patients with depression. More specifically, the research investigates whether interpretable, handcrafted acoustic feature-based methods, previously used for automatic speech-based depression detection, can be applied to detect depression in PD patients. We approach this by conducting a comparative study of speech-based depression detection tasks, using the DAIC-Woz corpus for typical speech and the PD-Depression corpus for atypical speech. We investigate how the acoustic descriptors of typical depressed speech differ from those extracted from the speech of PD patients suffering from depression. Our finding indicates that while typical depressed speech exhibits pronounced fluctuations in pitch and vocal stability, the speech of Parkinson's patients presents a more varied range of spectral features, reflecting the complexities of their condition caused by hypokinetic dysarthria.**

***Index Terms*—Parkinson's Disease, Depression detection, Acoustic features, Speech in health**

I. Introduction

Depressive disorder is a prevalent and complex mental illness that, according to the World Health Organization, afflicts approximately 5% of the global adult population [1]. The disorder is characterized by emotional, cognitive, and behavioral symptoms [2], including feelings of low mood or anxiousness, psychomotor retardation, or, in the worst cases, it can lead to suicide [3]. Assessment of this condition primarily relies on clinical interviews, which can lead to variability and heterogeneity in diagnosis [4]–[6]. Hence, there is a need to focus on the development of robust automated methods for screening depression in its various manifestations.

Within the speech community, researchers have conducted numerous studies utilizing acoustic features for detecting depression, showcasing the feasibility of vocal features [7]–[12]. These investigations have incorporated statistical analyses of acoustic descriptors such as fundamental frequency (F0), intensity, and spectral characteristics, indicating that these acoustic descriptors could potentially serve as effective indicators for depression screening. Over the last few years, Artificial Neural Network (ANN) based methods have gained widespread adoption within the community for the task of depression detection, showing great potential in analyzing speech patterns that are associated with the illness [13]–[15]. Despite generally achieving high accuracy compared to traditional handcrafted feature-based strategies, these approaches often sacrifice interpretability for the majority of models. Furthermore, they face challenges associated with limited training data.

Despite significant progress in the field, speech-based depression detection remains a challenging task, primarily attributed to the diverse manifestations of depression. A noteworthy aspect overlooked by many depression studies is the critical exploration of detecting depression in conjunction with other neurological disorders like Alzheimer's disease [16] or Parkinson's disease. Aarsland et al. [17] underlined as PD is predominantly known for its movement-related symptoms, but the disorder also involves a range of non-motor symptoms, with depression being the most prevalent, affecting roughly one-third of individuals with PD. It is often long-lasting and in some cases appears in prodromal states [18]

The coexistence of Depression and neurological disorders may introduce complexity to the diagnostic process, often resulting in the oversight and neglect of treatment for numerous cases [19], prompting notable concern. There is a significant gap in the existing literature, which lacks exploration into the detection of depression from an atypical speech. To the best of our knowledge, only three existing studies have addressed this specific case. Ozkanca et al. [20] were the first to conduct depression screening on PD patients. They recorded 10s phoneme sound samples from the subjects. Classification was performed by employing various machine learning classifiers, utilizing handcrafted features. Their findings demonstrated a strong correlation between voice and depression in PD. Pérez-Toro et.al [21]–[23] reported automatic speech-based depression detection on monologues for Parkinson's and Alzheimer's patients. They utilized transfer learning techniques, thereby modeling emotions on the valence-arousal plane using Forest-Net [24] and then finetuning for depression in AD and PD patients, highlighting the potential of leveraging emotional information for such diagnostic tasks. While [20] explored handcrafted features, the study is limited to phonemes. In

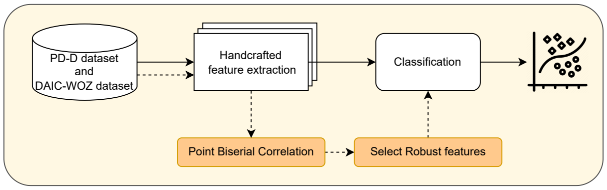

Fig. 1. Proposed methodologies representation: conventional approach (solid arrows) and PBCC-based approach (dashed arrows).

contrast, [21], [23] investigated monologues, but their work centered around emotion as the base task.

In this study, we ask the question of whether interpretable handcrafted feature-based methods, previously employed for automatic speech-based depression detection, can be extended to detect depression in PD patients. Additionally, we aim to analyze the acoustic features differences in a normal speech[1] and PD patients speech and what may pose challenges for the classifier in detecting depression. We do so by conducting a systematic analysis based on acoustic features, to compare non-PD patients with depression and PD patients diagnosed with depression. For this study, we utilize interpretable handcrafted feature representation as input to the classifiers. Additionally, we propose a method that utilizes the Point Biserial Correlation Coefficient (PBCC) for feature selection, to identify the most significant features for depression detection within the datasets which may help enhance the classification performance. Finally, for better understanding, we present our analysis by grouping acoustic features into three broad categories based on the type of information they convey: vocal source-related features describe the characteristics of the sound that affect pitch, loudness, and voice quality; vocal tract-related features describe the characteristics of the sound that speech articulator modifies as it travels through the vocal tract, such as formants; and global-related features describe both processes that influence speech production, such as energy.

The rest of the paper is organized as follows, Section II introduces the methods investigated in this study. Section III-A outlines the dataset used and the experimental setup, Section IV presents the experimental results and the analysis, and finally Section V concludes the paper.

II. Methodology

Figure 1 depicts the methodologies adopted for the study.

A. Conventional approach

Handcrafted features were first extracted from the input audio signal and then used as representations that are fed into the classifier block to generate the confidence score (Fig 1 solid arrows). This method serves as a valuable baseline for comparison with more complex strategies.

[1]In the context of this study, 'normal speech' refers to non-PD patient's speech.

B. Feature selection approach

Handcrafted features are first extracted from the raw audio, followed by feature selection using the Point Biserial Correlation Coefficient (PBCC) (Fig 1 dashed arrows). The PBCC quantifies the strength of the linear relationship between a binary target variable (D/ND) and continuous features, allowing for the identification of a smaller, more relevant subset of features. The measure is defined as:

$$r_{pb} = \frac{M_D - M_{ND}}{s_n}\sqrt{\frac{n_D n_{ND}}{n^2}}$$

where M_D and M_{ND} are the mean of the features for Depressive (D) and Non-Depressive classes respectively; n_D and n_{ND} represent the sample sizes of the two groups; n is the total amount of samples; s_n is the standard deviation of the continuous variables.

Initially, correlation threshold values are established. For each threshold, the PBCC is computed between each feature and the target labels of the training samples. Features with PBCC values that surpass the defined threshold are retained, as they exhibit a stronger linear relationship with the target variable and are deemed more informative. Subsequently, a GB Classifier is utilized to assess the predictive performance of the selected features. Different thresholds, ranging from 0.06 to 0.6 in increments of 0.2, are tested, and the subset of features that optimizes the system's accuracy is ultimately selected.

III. Experimental setup

A. Datasets and protocols

(a) Distress analysis interview corpus - Wizard of Oz (DAIC-WOZ) [25]: consists of audio-visual interviews from 193 participants assessed for psychological distress, including anxiety, depression, and PTSD, totaling 17 hours of audio data. Standard train-test split provided in the database, following the guidelines of the AVEC 2016 challenge [26]. Precisely, recordings from 107 interviews were utilized for training, and the evaluation was performed using recordings from 35 interviews, serving as the test set.

(b) Depression in Parkinson's disease (PD-D) [21]: consists of speech data from 60 Spanish speakers from Colombia, including 25 Depressive PD patients (D-PD) and 35 Non-Depressive PD patients (ND-PD). All participants were instructed to provide a monologue about their daily routines, after which a neurologist evaluated their neurological state using the Movement Disorders Society – Unified Parkinson's Disease Rating Scale (MDS–UPDRS) [27]. This scale is considered the standard for evaluating the neurological status of PD patients. The first part of the MDS-UPDRS scale contains an item that assesses depression based on the patient's daily routines, with scores ranging from 0 to 4. Patients with scores higher than zero were labeled as depressed PD patients denoted by D-PD, while those with scores equal to zero were classified as Non-Depressed PD patients indicated by ND-PD. The average duration of the monologues is 84±34 seconds for the D-PD patients and

TABLE I
DISTRIBUTION OF UTTERANCES USED IN THE STUDY, CORRESPONDING TO EACH LABEL.

Database	Content	Depressed patients	Not-Depressed patients	Total
DAIC-WOZ	English	42	100	142
PD-D	Spanish	24	35	59

80±37 for the ND-PD patients, for a total duration of about 4892 seconds for the whole dataset. For our study, we excluded speaker 52 due to recording errors. To be consistent with the previous work [21], we opted for the Leave One Speaker Out (LOSO) cross-validation protocol. That is, for evaluating the 'k'-th speaker the classifier was trained on the remaining 'k-1' speakers.

Table I summarizes the two datasets.

B. Features and classifiers description

The study employed three well-known sets of knowledge-based handcrafted features: EGEMAPS [28], COMPARE [29], both extracted using the openSMILE toolkit [30], and $DisVoice$ [31]. EGEMAPS includes 25 Low-Level Descriptors (LLDs) designed to capture key aspects of audio signals, such as frequency-related information, energy, amplitude, and spectral parameters. From this set of descriptors, a range of statistical functionals were computed, resulting in a comprehensive total of 88 distinct parameters. COMPARE offers a more extensive set with 6373 features calculated using delta functions for frame-level representation and functionals for utterance-level representation. Lastly, for $DisVoice$ feature, the static representations of phonation, articulation, and prosody features were combined to create a single representation. More details on the features and extraction code can be found in [31].

In this study, we utilized three distinct machine-learning algorithms for the classification task, namely Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting (GB). The optimal hyperparameter configuration for each classifier was selected based on the Grid Search along with 6-fold cross-validation performance, ensuring a robust model selection process.

C. Evaluation metrics

Following prior research [15], [32] we assessed the performance of our systems using *F1-score, precision,* and *recall* as the primary evaluation metrics. We presented the performance metrics for each class, depressed (D) and non-depressed (ND), for both datasets in the study. Additionally, we reported the overall (O) score as the unweighted average across the two groups (D and ND).

IV. RESULTS AND ANALYSIS

A. System performance

Table II presents the results for the best-performing classifier (GB) for each feature set across both corpora used. In the DAIC-WOZ dataset, our system significantly outperforms the baseline values reported by AVEC 2016, which were obtained using the EGEMAPS feature set with an SVM classifier, demonstrating an improvement of approximately 51%. Initially, EGEMAPS achieves a F1 score of 0.74 in the conventional approach; however, following feature selection, this score slightly decline to 0.69, likely due to its lower dimensionality. In contrast, both COMPARE and $DisVoice$ show notable enhancements in their performance after feature selection.

In the context of the PD-D corpus, our methodologies surpass the overall F1 score of 0.7 reported by Perez-Toro, who utilized a combination of the Valence and Arousal representation for classification [21]. Initially, $DisVoice$ features demonstrate superior performance within the conventional pipeline, achieving an F1 score of 0.74. After applying Point Biserial-based feature selection, we observe a significant performance improvement across all feature sets. Notably, COMPARE improves from 0.43 to 0.78, while utilizing only 186 features for classification. Overall, it is worth noting that feature selection enhances the classifier's ability to better predict depressed (D) patients in both corpora, the exception being EGEMAPS for DAIC-WOZ, as reflected in the F1 scores for the D column. These results show that the PBCC method filters out redundant features, leading to simpler, more interpretable models with improved precision-recall balance.

TABLE II
CLASSIFIERS' PERFORMANCE OVER THE TWO DATASETS. *Dims* DENOTES THE FEATURE DIMENSION; *Thr.* SIGNIFIES THE THRESHOLD SET FOR FEATURE SELECTION; *D* AND *ND* DENOTE DEPRESSED AND NOT-DEPRESSED PATIENTS, RESPECTIVELY; *O* IS THE UNWEIGHTED AVERAGE OF *D* AND *ND*.

			F1-score			Precision		Recall	
Features	Dims	Thr.	O	D	ND	D	ND	D	ND
DAIC-Woz									
Valstar et al. [26]	88		0.49	0.41	0.58	0.26	0.94	0.88	0.42
Conventional approach									
EGEMAPS	88		**0.74**	0.62	0.87	0.90	0.78	0.47	0.97
COMPARE	6373		0.47	0.24	0.07	0.45	0.59	0.16	0.86
$DisVoice$	620		0.55	0.33	0.77	0.50	0.69	0.25	0.87
Feature-selection approach									
EGEMAPS	39	0.18	0.69	0.54	0.84	0.67	0.72	0.33	0.91
COMPARE	2756	0.18	**0.72**	0.65	0.80	0.62	0.78	0.33	0.87
$DisVoice$	184	0.20	0.65	0.47	0.83	0.80	0.73	0.33	0.96
PD-D									
Perez-Toro et al. [21]			0.68	-	-	-	-	-	-
Conventional approach									
EGEMAPS	88		0.54	0.40	0.69	0.53	0.61	0.32	0.79
COMPARE	6373		0.43	0.30	0.56	0.33	0.53	0.28	0.59
$DisVoice$	620		**0.74**	0.69	0.78	0.71	0.77	0.68	0.79
Feature-selection approach									
EGEMAPS	7	0.26	0.65	0.57	0.72	0.62	0.68	0.52	0.76
COMPARE	186	0.3	**0.78**	0.75	0.81	0.73	0.82	0.76	0.79
$DisVoice$	16	0.32	0.77	0.73	0.81	0.75	0.80	0.72	0.82

B. Feature Analysis

Following the classification outcomes discussed in the previous section, we conducted a feature analysis by ranking the top 10 features based on the normalized feature importance values assigned by the best-performing classifier based on the feature selection approach for both databases. To gain a better understanding, we categorize the acoustic descriptors into three groups, considering the information conveyed by each feature. Using the classification from Eyben et al. [33] and, we distinguished Low-Level Descriptors into vocal source-related and vocal tract-related categories. Additionally, we introduced a third category for features that represent global speech signal information, incorporating both the vocal tract and source. Table III presents the features sorted by their importance scores using the GB model. The descriptor names are included for clarity. The "Index" column indicates the 'i'-th element from the indexed feature list in the openSMILE header, while the "Group" column shows our proposed categories.

TABLE III
FEATURE RANKING OF GB TRAINED ON COMPARE FOR BOTH DAIC-WOZ (LEFT) AND PD-D (RIGHT), USING PBCC FEATURE SELECTION APPROACH.

DAIC-WOZ			PD-D		
Index	**LLD name**	**Group**	**Index**	**LLD name**	**Group**
3861	logHNR	Source	142	Length L1	Global
4038	Jitter DDP	Source	88	RMS Energy	Global
5126	Py Sharpness	Source	1	Length L1	Global
1493	Spectral Harmonicity	Source	64	RMS Energy	Global
6077	Spectral Variance	Global	6067	Spectral Entropy	Vocal tract
6174	MFCC	Vocal tract	1245	Spectral Flux	Global
2365	audSpec	Vocal tract	1476	Py Sharpness	Source
4132	F0	Source	2925	Spectral RollOff 90.0	Vocal tract
4131	F0	Source	2991	Spectral Centroid	Global
1373	Spectral Skeweness	Vocal tract	3106	Spectral Kurtosis	Global

The feature rankings displayed in Table III highlight a significant difference in how the classifier detects depression in non-PD versus PD speech. In the case of the DAIC-WOZ corpus, where the classification involves distinguishing between depressive patients and healthy control subjects, the classifier predominantly focuses on source-related features (6 out of 10), a select number of vocal tract features (3 out of 10), and a single global-related low-level descriptor (LLD). This emphasis on source features indicates that characteristics reflecting the quality and stability of vocal fold vibrations—such as harmonic-to-noise ratio, jitter, and pitch variations—are crucial for detecting depression within the non-pathological population represented in the DAIC dataset. Additionally, these findings are consistent with observations made by several clinicians treating patients with depression. Hollien [34] notes alterations in pitch patterns among depressed patients, while Darby [35] highlights a reduction in pitch and vocal intensity. In contrast, the PD-D dataset the feature ranking reveals a more diverse distribution of information across various features. The inclusion of multiple spectral features, such as *spectral entropy*, *spectral centroid*, *spectral roll-off*, *spectral flux*, and *spectral kurtosis*, underscores the instability of vocal characteristics in patients with Parkinson's disease [36]. Furthermore, the noted reductions in *spectral entropy* and *spectral centroid* among individuals experiencing depression illustrate the complex relationship between emotional states and voice quality [37]. The *Length L1*, which serves as a measure of voice quality and stability, is significantly affected, indicating fluctuations in speech patterns that may stem from motor control issues associated with the disease. This shift emphasizes the complexities introduced by motor control issues and vocal tract coordination deficits in analyzing speech for depression detection in individuals with PD.

V. CONCLUSION

This study explores automatic depression detection in PD patients using the PD-D corpus and compares it with typical depression-affected continuous speech using the DAIC-WOZ corpus. The classification findings suggest that hand-crafted feature-based methods for detecting depression in patients with Parkinson's Disease can be extended. The performance of the Gradient Booster classifier shows a significant improvement when employing the Point Biserial Correlation-based feature selection approach rather than the conventional method across both datasets. This result underscores the importance of eliminating redundant features to develop more robust models.
The findings from the analysis of feature rankings reveal the difference in terms of acoustic descriptors of depression in non-PD speech versus PD speech. For DAIC-WOZ dataset the classifier emphasizes source-related features and demonstrates their crucial importance in identifying depressive states. This observation is supported by existing research, which underlines that fluctuations in vocal characteristics, such as pitch variations, serve as key indicators of depression. Conversely, in PD speech, a broader range of spectral features stands out, capturing the complexities introduced by Parkinson's disease manifestations. These characteristics are likely closely linked to speech motor symptoms, which are prevalent among Parkinson's disease patients. Our findings indicate that the symptoms of Parkinson's make it challenging for the classifier to classify depression automatically. The study highlights that depression, along with Parkinson's disease, can present distinct acoustic characteristics. These findings may contribute to advancing our understanding of speech analysis in the context of mental health in patients with neurodegenerative disorders.

ACKNOWLEDGMENT

This work was partially funded by (a) the SNSF through the Bridge Discovery project EMIL: Emotion in the loop - a step towards a comprehensive closed-loop deep brain stimulation in Parkinson's disease (grant no. $40B2-0\_194794$), and (b) CODI at UdeA, grant # PI2023-58010.

REFERENCES

[1] World Health Organization "Depression", https://www.who.int/news-room/fact-sheets/detail/depression, last accessed 19/10/2023
[2] Scibelli, F., Roffo, G., Tayarani, M., Bartoli, L., Mattia, G., Esposito, A., Vinciarelli, A.: Depression Speaks: Automatic Discrimination between Depressed and Non-Depressed Speakers Based on Nonverbal Speech Features. *Proc. of ICASSP* (2018)
[3] Global Health Estimates: Life expectancy and leading causes of death and disability, https://www.who.int/health-topics/mental-health, last accessed 15/01/2025.

[4] Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., Quatieri, T.: A review of depression and suicide risk assessment using speech analysis. *Speech Communication* (2015)

[5] Landau, M.: Acoustical Properties of Speech as Indicators of Depression and Suicidal Risk. *Vanderbilt Undergraduate Research Journal* (2008)

[6] Burdick, B., Holmes, C., Waln, R.: Recognition of suicide signs by physicians in different areas of specialization. *Journal Of Medical Education* (1983)

[7] Nilsonne, ėt al.: Measuring the rate of change of voice fundamental frequency in fluent speech during mental depression. *The Journal of the Acoustical Society of America*. (1988)

[8] France, D. et al.: Acoustical properties of speech as indicators of depression and suicidal risk. *IEEE Transactions On Biomedical Engineering*. (2000)

[9] Moore II, E. et al.: Critical analysis of the impact of glottal features in the classification of clinical depression in speech. *IEEE Transactions On Biomedical Engineering*. (2007)

[10] Cummins, N. et al.: An investigation of depressed speech detection: Features and normalization. *Proc. of INTERSPEECH*. (2011)

[11] Stasak, B. et al.: An investigation of emotional speech in depression classification. *Proc. of INTERSPEECH*. (2016)

[12] Zahid, L. et al.: Detection of Speech Impairments in Parkinson Disease Using Handcrafted Feature-Based Model on Spanish Speech Corpus. *Proc. of CCIS*. (2020)

[13] He, L., Cao, C. :Automated depression analysis using convolutional neural networks from speech. *J. Biomed. Inform.*. (2018)

[14] Ma, X. et al.: DepAudioNet: An efficient deep model for audio based depression classification. *Proc. of AVEC, On ACM MM*. (2016)

[15] Dubagunta, S., Vlasenko, B. & Doss, M.: Learning voice source related information for depression detection. *Proc. of ICASSP 2019*

[16] Lee, H., Lyketsos, C.: Depression in Alzheimer's disease: heterogeneity and related issues. *Biological Psychiatry*. (2003)

[17] Aarsland, D., Påhlhagen, S., Ballard, C., Ehrt, U. & Svenningsson, P.: Depression in Parkinson disease—epidemiology, mechanisms and management. *Nature Reviews Neurology*. (2012)

[18] Dušek, Petr, et al.: Relations of non-motor symptoms and dopamine transporter binding in REM sleep behavior disorder. *Nature Scientific reports*. (2019)

[19] Hussain, M. et al.: Similarities between depression and neurodegenerative diseases: pathophysiology, challenges in diagnosis and treatment options. *Cureus*. (2020)

[20] Ozkanca, Y., Göksu Öztürk, M., Ekmekci, M., Atkins, D., Demiroglu, C. & Hosseini Ghomi, R.: Depression screening from voice samples of patients affected by parkinson's disease. *Digital Biomarkers* (2019)

[21] Pérez-Toro, P., Vásquez-Correa, J., Bocklet, T., Nöth, E. & Orozco-Arroyave, J.: User state modeling based on the arousal-valence plane: Applications in customer satisfaction and health-care. *IEEE Transactions On Affective Computing* (2021)

[22] Pérez-Toro, Paula Andrea, et al.: Depression assessment in people with Parkinson's disease: The combination of acoustic features and natural language processing. *Speech Communication* (2022)

[23] Pérez-Toro, P. et al.: Transferring Quantified Emotion Knowledge for the Detection of Depression in Alzheimer's Disease Using Forestnets. *Proc. of ICASSP*. (2023)

[24] Rodríguez-Salas, D., et al.: Mapping ensembles of trees to sparse, interpretable multilayer perceptron networks. *SN Computer Science*. (2020)

[25] Gratch, J., Artstein, R., Lucas, G., Stratou, G., Scherer, S., Nazarian, A., Wood, R., Boberg, J., DeVault, D., Marsella, S. & Others: The distress analysis interview corpus of human and computer interviews. *LREC* (2014)

[26] Valstar, M., Gratch, J., Schuller, B., Ringeval, F., Lalanne, D., Torres Torres, M., Scherer, S., Stratou, G., Cowie, R. & Pantic, M.: Avec 2016: Depression, mood, and emotion recognition workshop and challenge. *AVEC* (2016)

[27] Goetz, C., Tilley, B., Shaftman, S., Stebbins, G., Fahn, S., Martinez-Martin, P., Poewe, W., Sampaio, C., Stern, M., Dodel, R. & Others: Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Movement Disorders: Official Journal Of The Movement Disorder Society* (2008)

[28] Eyben, F., Scherer, K., Schuller, B., Sundberg, J., André, E., Busso, C., Devillers, L., Epps, J., Laukka, P., Narayanan, S. & Others: The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Transactions On Affective Computing* (2015)

[29] Schuller, B., Steidl, S., Batliner, A., Vinciarelli, A., Scherer, K., Ringeval, F., Chetouani, M., Weninger, F., Eyben, F., Marchi, E. & Others: The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism. *Proc. of INTERSPEECH* (2013)

[30] Eyben, F., Wöllmer, M. & Schuller, B.: Opensmile: the munich versatile and fast open-source audio feature extractor. *Proceedings Of The 18th ACM International Conference On Multimedia* (2010)

[31] Orozco-Arroyave, J., Vásquez-Correa, J., Vargas-Bonilla, J., Arora, R., Dehak, N., Nidadavolu, P., Christensen, H., Rudzicz, F., Yancheva, M., Chinaei, H. et al.: NeuroSpeech: An open-source software for Parkinson's speech analysis. *Digital Signal Processing* (2018)

[32] Rinaldi, A., Tree, J. & Chaturvedi, S.: Predicting depression in screening interviews from latent categorization of interview prompts. *Proceedings Of The 58th Annual Meeting Of ACL* (2020)

[33] Eyben, F. & Eyben, F.: Acoustic features and modelling. *Real-time Speech And Music Classification By Large Audio Feature Space Extraction* (2016)

[34] Hollien, H.: Vocal indicators of psychological stress. *Annals Of The New York Academy Of Sciences* (1980)

[35] Darby, J., Simmons, N. & Berger, P.: Speech and voice parameters of depression: A pilot study. *Journal Of Communication Disorders* (1984)

[36] Hauptman, Y., Aloni-Lavi, R., Lapidot, I., Gurevich, T., Manor, Y., Naor, S., Diamant, N. & Opher, I.: Identifying Distinctive Acoustic and Spectral Features in Parkinson's Disease. *Proc. of INTERSPEECH* (2019)

[37] Hussenbocus, A. et al.: Statistical differences in speech acoustics of major depressed and non-depressed adolescents. *Proc. of ICSPCS*. (2015)