INTERSPEECH 2019: Show & Tell Contribution
September 15–19, 2019, Graz, Austria

Depression State Assessment: Application for detection of depression by speech

Gábor Kiss, Dávid Sztahó, Klára Vicsi

Budapest University of Technology and Economics, Budapest, Hungary
gabor.kiss@tmit.bme.hu, sztaho@tmit.bme.hu, vicsi@tmit.bme.hu

Abstract

We present an application that detects depression by speech based on a speech feature extraction engine. The input of the application is a read speech sample and the output is predicted depression severity level (Beck Depression Inventory). The application analyses the speech sample and evaluates it using support vector regression (SVR). The developed system could assist general medical staff if no specialist is present to aid the diagnosis. If there is a suspicion that the speaker is suffering from depression, it is inevitable to seek special medical assistance. The application supports five native languages: English, French, German, Hungarian and Italian.

Index Terms: depression, SVR, speech, support vector regression, pathological speech, voice analysis, voice disorder

1. Introduction

Depressed state affects human speech, such as speech dynamics, speech clarity, and speech articulation. Thus speech is a promising indicator as a biomarker for depression detection.

An automatic prediction system will be presented, which can predict the severity of the depression of examined subjects based on their speech signal. This work has been started under an ESA project, titled the "Psychological Status Monitoring by Computerized Analysis of Language phenomena (COALA-Phonetics)" and continued under the Project no. K128568 of National Research, Development and Innovation Fund of Hungary.

Depression is a psychiatric disorder. Several events can cause depressed state in the life of a person, like stressful events, persistent sadness, difficulties in the daily duties, lack of sunshine and isolation from the outside world, and so on [1]. Upon the development of depression, the depressed person's quality of life and ability to work are impaired, and there is an increased chance of suicide. The World Health Organization (WHO) Depression - the publication of Global Public Health Concern - stated that there were 350 million people suffering from depression in 2012 [2]. The WHO predicts that by 2030 unipolar depression will be among the top three most serious diseases worldwide, alongside HIV/AIDS and heart disease [3].

Precise diagnosis can only be made by medical specialist, but in some cases this is not feasible. This was the case at Concordia research station at Antarctica, where some researchers stayed for a whole year. Despite the lack of medical specialist, it was necessary to monitor the psychological status of the researchers, especially the possibility of developing depression. The current application, Depression State Assessment, was developed for this purpose, which is able to estimate the depressed condition of the speaker using statistical and machine learning regression methods based on speech processing. The application is not intended to use as a standalone clinical diagnostic tool. It can assist medical staff if no specialist is present to aid the diagnosis. If there is a suspicion that the speaker is suffering from depression, it is inevitable to seek medical assistance.

The application was developed using the Reference Depressed Speech Database, which contains speech samples from healthy and depressed speakers from three languages: 150 German [8], 127 Hungarian and 11 Italian speakers.

2. Overview of the application

The application was created for depression detection based on speech processing. The input of the application is a read speech sample and the output is a predicted BDI (Beck Depression Inventory [7]) score, which represents the severity of the depressed state of the speaker. The following partitioning is usually given for the range of BDI scores: 0-13 – minimal depression (healthy); 14-19 – mild depression; 20-28 – moderate depression; 29-63 – severe depression.

The application analyses the speech sample and evaluates it using support vector regression (SVR). The evaluation is done by a pre-trained SVR model. For the details on the training, acoustic features and achievable accuracy, see [4][5] and [6].

To use the application, the user has to read a predefined text (the tale, called 'The North Wind and the Sun') in his or her native language. The text is shown on the screen and the application supports five native languages:

- English,
- French,
- German,
- Hungarian,
- Italian.

The input speech sample is segmented and labelled into phoneme-like units. Each sound file is characterized by a feature vector containing preselected acoustic-phonetic parameters. The input of the automatic predictor is the pre-trained model (obtained using a training dataset) and the normalized feature vector. The output of the automatic prediction system is the predicted BDI score.

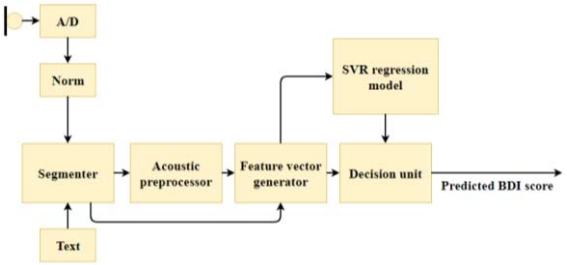

Figure 1.: *Process of decision making*

The dataflow diagram of the Depression State Assessment application can be seen in Figure 1. The parts are the following:

- **A/D**: performs analogue digital conversion and records speech.
- **Norm**: normalizes the amplitude of the speech to peak.
- **Segmenter**: segments the input speech sample into phoneme-like units using a forced alignment algorithm.
- **Acoustic preprocessor**: measures and calculates the acoustic and phonetic parameters based on the sex of the speaker. These parameters are statistically well-selected, quasi language-independent parameters.
- **Feature vector generator**: produces the appropriate feature vector from the measured acoustic and phonetic parameters valid for multiple languages and normalizes its values.
- **SVR model**: the model containing the statistical knowledge of the regression problem. It is derived from a training phase using Support Vector Regression method based on Reference Depressed Speech Database and it is used as one of the input of the decision unit.
- **Decision unit**: predicts the severity of depression in BDI using Support Vector Regression method based on the feature vector and the SVR regression model.
- **Testing the model**: The pre-trained model was tested with leave-one-out cross-validation method on the Reference Depressed Speech Database. The essence of the leave-one-out cross-validation method is that it always uses one sample for testing and the remainder samples for training, so the two sets are disjoint and every sample is used for testing.

Table 1 shows the main descriptive characteristics of the pre-trained model (mean absolute error and root mean square error), complemented by the Pearson correlation coefficient of the original and predicted BDI scores. There are no other language independent depression recognition researches that we are aware of. Therefore, no result comparison is included in the Tables. The user interface of the application is shown in Figure 2.

Table 1: *Descriptive characteristics of testing on Reference Depressed Speech Database*

MAE	RMSE	Pearson correlation
6.7	8.9	0.7

Table 2: *Results of the tests by native speakers.*

MAE	RMSE
7.3	9.1

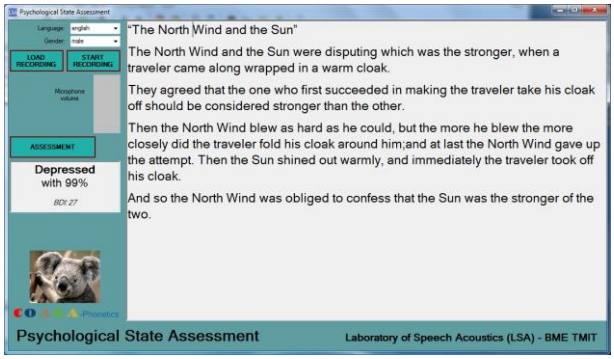

Figure 2: *User interface of the software*

3. Experience of use

The application was tested by three native speakers from each language. Table 2. shows the main descriptive characteristics of the test results (mean absolute error and root mean square error values).

4. Conclusions

An application was presented for detection of depression by speech based on the speech feature extraction engine. It could assist medical staff if no specialist is present to aid the diagnosis. The application analyses the speech sample, extract specific acoustic-phonetic features and evaluates it using support vector regression (SVR). Results of the pre-trained models and the first user experiences show that the application can be a promising tool for medicals or GPs. The performance of the model could be continuously increased by obtaining larger number of audio samples.

5. Acknowledgements

Project no. K128568 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the K_18 funding scheme.

The research was partly supported by European Space Agency COALA project: Psychological Status Monitoring by Computerized Analysis of Language phenomena (COALA) (AO-11-Concordia).

6. References

[1] Schumann, I., Schneider, A., Kantert, C., Lo¨we, B., Linde, K., (2012) Physicians' attitudes, diagnostic process and barriers regarding depression diagnosis in primary care: a systematic review of qualitative studies. *Fam. Pract*. 29, 255–263.

[2] Marcus, M., Yasamy, M. T., van Ommeren, M., Chisholm, D., & Saxena, S. (2012) Depression: A global public health concern. WHO Department of Mental Health and Substance Abuse, 1, 6-8.

[3] Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Breakspear, M., & Parker, G. (2013) Detecting depression: a comparison between spontaneous and read speech. IEEE Int. Conference on Acoustics, Speech and Signal Processing , 7547-7551

[4] Kiss, G., Sztahó, D., & Vicsi, K. (2013) Language independent automatic speech segmentation into phoneme-like units on the base of acoustic distinctive features. IEEE CogInfoCom 2013, 579-582

[5] Kiss, G., & Vicsi, K. (2017. Mono-and multi-lingual depression prediction based on speech processing. *Int. J. of Speech Technology*, 20(4), 919-935.

[6] Kiss, G., & Vicsi, K. (2017) Follow-up Psychological Status Monitoring of the Crew Members of Concordia Research Station at Antarctica Based on their Speech. 3rd Int. Conf. on Research, Technology and Education of Space (p. 37).

[7] A. T. Beck, R. A. Steer, R. Ball and W. F. Ranieri (1996) Comparison of beck depression inventories -IA and -II in psychiatric outpatients, *Journal of Personality Assessment* 67, 588–597

[8] Valstar, M., Schuller, B., Smith, K., Eyben, F., Jiang, B., Bilakhia, S., & Pantic, M. (2013) AVEC 2013: the continuous audio/visual emotion and depression recognition challenge. 3rd ACM Int. Workshop on Audio/visual emotion challenge, 3-10