Speech Analysis and Depression

Tan Tze Ern Shannon*, Dai Jingwen Annie† and See Swee Lan †
*Raffles Girls School (Secondary), 20 Anderson Road, 259978 Singapore
E-mail: shannon.tan00@student.rgs.sg Tel: +65 64082000
Email: annie.dai00@student.rgs.sg
†Institute for Infocomm Research (I2R), A*STAR, 1 Fusionopolis Way, Singapore 138632
E-mail: slsee@i2r.a-star.edu.sg Tel: +65 64082000

Abstract—In this paper, the correlation between the speech features of the vowel /a/ and depression severity was investigated, so as to derive a depression severity meter mobile application that can accurately detect depression quantitatively. Results showed a correlation between depression severity and speech features, and an application prototype was created and tested to assess for predictive accuracy of BDI score.

I. BACKGROUND

Depression is on the rise [1] and there is an increasing need for efficient and accurate diagnosis for effective treatment. With the increasing prevalence of smartphone usage [2], mobile applications can be considered as a method that can be utilised as an aid for psychiatrists. They are not only able to give immediate user feedback, but also able to reach out to users more easily than traditional therapy sessions due to their ubiquity, making monitoring of depression more convenient. These applications mainly make use of qualitative or quantitative analysis. Applications currently in the app stores such as CBT Self-Help Guide [3] and What’sMyM3 [4] utilise qualitative analysis, which can be subjected to bias, such as the social desirability bias wherein respondents may answer falsely to hide traits which they deem as undesirable. While research regarding the use of quantitative analysis to assess depression severity has been explored [5][6], current applications developed such as MoodTracer [7] still have their limitations. It measures depression severity in batch processing and the result is not available instantaneously. Therefore, there is still a need for an application that can provide automatic, immediate and accurate response in the detection of depression severity.

The aim of this research is to develop an application that is able to utilise quantitative analysis to assess depression severity, using biomarkers, or speech features. Literature review has found that vocal features are potential biomarkers in the detection of depression severity as they bear statistical correlations with depression across many studies [8].

In this paper, the correlations between depression severity and speech features are analysed, specifically formant frequencies and jitter, during the sustained phonation of the vowel /a/. Other features have also been investigated but these 2 features have been chosen as the focus for this paper. Formant frequencies have been loosely associated with vocal articulation and therefore is one means of which articulatory changes in a depressed voice can be assessed [9]. Jitter, defined as the vocal fold vibrations, can be used as a means to measure laryngeal control, a form of motor control [8]. Depression has been found to affect motor control, and it has been found that psychomotor retardation affects laryngeal biomarkers, causing changes in quality of speech. Therefore, they may possibly be ideal as biomarkers of depression.

In addition, the use of the sustained phonation of the vowel /a/ as a means to detect depression severity is much less time consuming than proposed methods of reading passages or free speech. It is also not limited by the language of the passages or of the application. Although this method has yet been employed in the detection of depression severity, it has found to be useful in the detection of Parkinson’s Disease (PD) since speech signal has been found to be useful in the diagnosis and detection of the disorder [10]. Researchers have found that the use of sustained phonation of the vowel /a/ allows for the detection of the severity of PD [11]. As prior research found that more females have depression [12], in this study, female adult patients were analysed to account for the biological differences between different age and gender groups [13].

II. METHODOLOGY

A. Speech Analysis

47 recordings of female adult depression patients saying the vowel /a/ for a sustained period of time of approximately 8 seconds were randomly extracted from the Audio/Visual Emotion recognition Challenge (AVEC) 2013 database [14]. They were randomly split into two groups, namely the analysis sample group, containing 24 recordings, and the test sample group, containing 23 recordings. The analysis sample group recordings were run through Praat [15], a speech analysis software, to obtain the jitter and formant frequencies values of the patients. Due to the different methods of computing jitter values in the software, Jitter (local, absolute) was selected as it is the absolute difference between consecutive seconds, as well as produces more precise results than the other measurements [16]. Formant frequencies 1 (F1), 2 (F2) and 3 (F3) were measured as these three are most commonly used to assess speech quality and articulation [17].

The values of the above four speech features were then statistically analysed for their correlations with depression severity, measured in Beck Depression Inventory (BDI) scores.

B. Depression Severity Meter Phone Application Prototype
Using the data generated from the above analysis, an algorithm to calculate user’s depression severity was constructed using the multiple linear regression model. The application prototype, HearMeOut, was developed using Android Studio, and was built with API 8 to target 100% of the Android devices active in Google Play. Android was chosen as the operating system as it is used in many different handset vendors and service providers [18], with a large portion of the world population using Android smartphones or tablets [19], allowing a wider potential reach to users.

Our application was then tested using the data obtained by running the recordings in the test sample group in Praat of female adult depression patients doing the same task, extracted from the same AVEC 2013 database. Predictive accuracy of the prototype was measured using Mean Absolute Error (MAE), whereby the difference in BDI score generated by the prototype was compared against that of the patient’s actual score. Root Mean Squared Error (RMSE) was also used to take into account large errors.

III. RESULTS

A. Speech Analysis

The Pearson product-moment correlation coefficient computation was used in the assessment between the speech features, jitter and formant frequencies, and the patient’s BDI score. A correlation of \( r > \pm 0.37 \) is considered strong, \( r > \pm 0.24 \) is moderate, \( r > \pm 0.10 \) is weak and negligible between 0.00 and \( \pm 0.09 \) [20].

It was found that there was a strong correlation between BDI score and jitter \( [r=0.47, n=24] \) (see Fig. 1) shows that as BDI score increases, jitter also increases.

The bivariate correlation between the BDI Score and F1 was positive at \( [r=0.37, n=24] \) (see Fig. 2). For F2, the bivariate correlation between the BDI Score and F2 was positive at \( [r=0.32, n=24] \) (see Fig. 3). Lastly, for F3, the bivariate correlation between the BDI Score and F3 was negligible \( [r=-0.14, n=24] \) (see Fig. 4). This shows that while there is a strong and moderate correlation between F1 and F2, and BDI score respectively, there is weak correlation between BDI score and F3. Hence, it can be concluded that as BDI score increases, the F1 and F2 may increase for the vowel /ɑ/, but the F3 will most likely be unaffected.

B. Depression Severity Meter Phone Application Prototype

From the data processed above, jitter, F1 and F2 frequencies were chosen as the biomarkers to be implemented into our prototype, as they are moderately, or highly correlated with depression severity.

The prototype first shows a homepage to introduce users to the application. In the next interface, users would be prompted to record themselves saying the vowel /ɑ/ by pressing the “Start Recording” button (see Fig. 6). The “Stop Recording” button would henceforth stop the recording and save the recording file into the mobile device’s files.

By pressing the “Next” button, users will be directed to the next interface (see Fig. 7) where they will click on a link that will bring them to a website (see Fig. 8). The website will instruct them to download the relevant software on their computers, as well as a Praat Script that allows them to easily
obtain the values of the speech features required. Praat was not implemented into the prototype due to copyright issues. Such a roundabout way was hence introduced in order for the jitter, F1 and F2 values to be generated.

After downloading the necessary software, users will return to the application and by clicking the “Next” button, an interface that prompt users to input the jitter and F1 and F2 values will follow (see Fig. 9). By pressing the “Generate Score” button, the application will then generate the BDI score using an algorithm and generate a statement to reflect the score (see Fig. 10).

The algorithm to generate the BDI score was created using multiple linear regression analysis:

\[
\text{BDI} = -24.57 + 0.028(F1) + 0.0064(F2) + 0.38(Jitter)
\]  \hspace{1cm} (1)

If the BDI score generated is above the score of 63, the limit of the BDI range, the application will generate the statement “Your BDI score could not be generated”. In the case where the BDI score is below 0, it will be assumed the user to be not depressed and generate the statement “You are not depressed”. The MAE and the RMSE of our model was 8.734 and 9.670 respectively, which are promising results suggesting that this application has much potential in the detection of depression severity in female adults.

IV. DISCUSSION

A. Speech Analysis

The high correlation (see Fig. 1) between BDI scores and jitter could be attributed to the irregular vocal fold vibrations from psychomotor retardation, a key aspect and symptom of depression [21]. Psychomotor retardation is loosely defined as any slowing (retardation) during a depressed state [22]. Past research also supports that jitter values are higher in depressed speech [23]. Hence, jitter is a suitable biomarker for assessing depression severity.

Formants are widely used in research relating to affective computing [24]. It is suggested that depression causes the tightening of the vocal tract due to psychomotor retardation, resulting in air having to vibrate at a higher rate in the cavity and thus higher formant frequencies [25]. From our data (see Fig. 2, Fig. 3 and Fig. 4), while moderate correlations between BDI values, and F1 and F2 are shown, there is a weak correlation when F3 was measured. This could be attributed to the fact that the vowel /ɑ/ requires more use of the pharyngeal and oral cavity, as compared to the nasal cavity. Therefore F1 and F2, which measure the vibration of air in the pharyngeal and oral cavity respectively, have stronger correlations than F3, which measure that air vibration in the nasal cavity.

B. Depression Severity Meter Phone Application Prototype

The MAE and the RMSE of our model was 8.734 and 9.670 respectively, suggesting that slight inaccuracies are present. Inaccuracies could stem from the fact that given only three variables were chosen to generate the BDI score; such a number was not substantial enough to eliminate significant errors (eg. a user may have naturally high jitter), skewing the supposed BDI score. In addition, the small sample size used for the generation of the algorithm may have resulted in inaccuracies in the prediction. However, with more refinement of the algorithm and a larger group of variables to eliminate significant errors, as well as a bigger sample size, it can potentially be an effective tool that can aid psychiatrists in
the monitoring and treatment of depression. The prototype shows much promise, and could potentially be the foundation of which an accurate and effective application can be designed to detect depression severity.

V. CONCLUSION

Our research has found that jitter, F1 and F2 are suitable biomarkers for the detection of depression severity due to the correlation between these speech features and depression severity. These findings are crucial in the prototyping of an automatic means of measuring depression severity.

Other suitable models could be looked into to be adapted into the system so as to increase the predictive accuracy. Increasing user friendliness by incorporating other methods of generating jitter and formant frequencies values into the system directly could be explored.

Ultimately though a larger sample size for testing would be ideal, all research done thus far can and will contribute to the end goal of developing a depression severity meter application. Such a tool is able to predict the user’s depression severity as accurately as possible, potentially being a helpful aid for psychiatrists in the monitoring and treatment of depression.

ACKNOWLEDGMENT

We would like to thank the Ministry of Education, Singapore and the Institute for Infocomm Research, A*STAR for providing us with this opportunity to research on this Science Mentorship Programme (SMP) project.

REFERENCES


