

PROPOSAL OF A DEPRESSION DETECTOR

Wei Tong Mok¹, Rachael Sing¹, Xiuting Jiang¹, Swee Lan See²

¹ Raffles Girls School (Secondary), 20 Anderson Road, Singapore 269978

² Institute for Infocomm Research (I²R), A*STAR, 1 Fusionopolis Way, Singapore 138632

ABSTRACT

Rapid advancements in technology, coupled with the increasingly widespread usage of social media, have brought about various impacts on our psychological health. Past research has been conducted on Singaporeans and ascertained the relationship between frequency of social media usage and depression levels in teenage girls. This research aims to investigate the relationship between depression and voice characteristics namely pitch, loudness and number of pauses, so as to determine markers for depression. The markers were then incorporated into a smartphone game designed to detect depression level through voice analysis. The likelihood of an individual suffering from depression was found to increase with loudness and frequency of pauses and to decrease with the frequency of an individual's voice. A mock up of the smartphone application was created on *Justinmind* and *Powerpoint*.

Index Terms: Social media, depression, voice characteristics

I. INTRODUCTION

Social media has become an indispensable part of almost everyone's life. According to O'Keeffe and Clarke-Pearson [1], social media are sites with networking, gaming and virtual world elements that allow social interaction. While it has brought about many positive impacts by increasing quality of life substantially and bringing about greater efficiency, it has also resulted in negative effects that should not be overlooked.

Past research has shown the strong correlation between frequency of social media and depression among Singaporean teenage girls aged 13-19 [2]. With the increasing number of social media users in Singapore, such a correlation would mean a growing number of teenage girls at risk of depression. This can pose a serious threat to the society because depression, a social challenge of psychological health concerns, burdens the social and economic development of our society in high healthcare utilization and costs.

Apart from their more sensitive nature and response to social-media induced stress, female teenagers are at a high risk of depression because they do not receive the guidance needed as they start to face symptoms. Rather, some of them turn to risky

internet sites and blogs for "help" that may promote behaviors that worsen their condition [2]. Therefore, this research aims to propose a smartphone application targeted at teenage girls to detect depression and provide reliable advice so that early treatment can be sought.

The application was designed as a maze, whereby the player would have to read a given passage to navigate, as his/her voice was recorded. The recording would then be analysed in a software called *Praat*, for voice features. Verbal reaction time would also be assessed. The results were then compared to depression markers to determine the player's depression level. Prior to constructing the application, videos of patients with varying depression levels were analysed for their voice characteristics namely frequency, loudness and number of pauses, to derive the depression markers.

II. METHODOLOGY

To investigate the relationship between voice characteristics and depression so as to determine depression markers, the speech of 12-13 patients with different Beck Depression Inventory (BDI) depression levels [3] were analysed for pitch/Hz, loudness/dB and number of pauses using *Praat*.

The BDI depression levels that were analysed are as follows,

- 0–13: minimal depression
- 14–19: mild depression
- 20–28: moderate depression
- 29–63: severe depression.

The game, called *A-maze-ing*, was designed as a maze, mocked-up on *Powerpoint* and *Justinmind*. To navigate within the maze, the player would have to read a given passage, as his/her voice was recorded. The recording would then be analysed in a software called *Praat*, for voice features. Additionally, there would be a junction where the player would be required to read the word flashed aloud and as quickly as possible to record verbal reaction time.

III. RESULTS

A. Video Analysis

From the analysis of the speech of patients with varying BDI depression levels, the results obtained were used to calculate the size and direction of the

linear relationship between depression levels and loudness, frequency and number of pauses. A bivariate correlation test was conducted. Graphs were then plotted to determine markers for the depression levels.

| | | Depression Level | Loudness |
|------------------|-----------------------------|------------------|----------|
| Depression Level | Pearson | 1.00 | .32 |
| | Correlation Sig. (2-tailed) | | .31 |
| | N | 12 | 12 |
| Voice Loudness | Pearson | .32 | 1.00 |
| | Correlation Sig. (2-tailed) | .31 | |
| | N | 12 | 12 |

Figure 1. Correlation between depression level and loudness

As shown in Figure 1, the bivariate correlation between the level of depression and average voice loudness was moderate, $r(10) = 0.32$. 10.24% of the variability in the patients' average voice loudness can be accounted for by their depression level. The positive coefficient indicates that as depression level increases, loudness increases as well (see Figure 2).

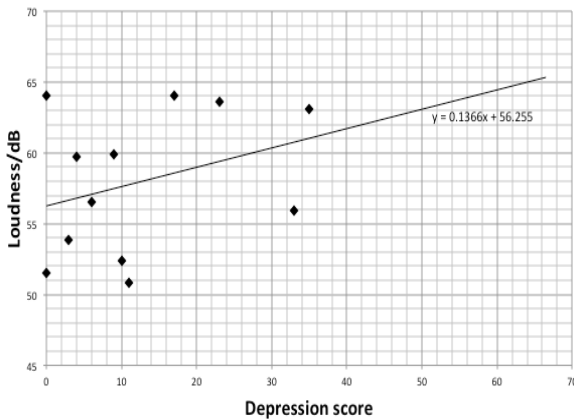


Figure 2. Graph showing relationship of depression score against loudness/dB

| | | Depression Level | Frequency |
|------------------|-----------------------------|------------------|-----------|
| Depression Level | Pearson | 1.00 | -.31 |
| | Correlation Sig. (2-tailed) | | .30 |
| | N | 13 | 13 |
| Voice Frequency | Pearson | -.31 | 1.00 |
| | Correlation Sig. (2-tailed) | .30 | |
| | N | 13 | 13 |

Figure 3. Correlation between depression level and voice frequency

As shown in Figure 3, the bivariate correlation between the depression level and average voice frequency was moderate, $r(11) = -0.31$. 9.61% of the variability in the patients' average voice frequency

can be accounted for by their depression level. The negative coefficient indicates that as depression level increases, frequency decreases (see Figure 4).

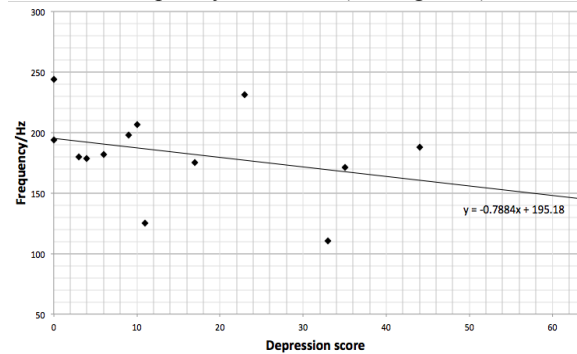


Figure 4. Graph showing relationship between Depression score against voice frequency/Hz

The table below shows the markers derived from the depression videos. The baseline for having depression would be the mild depression level, with frequency at a low of 184.136Hz and loudness at a high of 58.167dB. The baseline is as mild depression rather than minimal/no depression because it is assumed that intervention would only be necessary for individuals with mild depression.

From the equation of the two graphs (Figures 2 and 4), the markers for depressing depression can be concluded as in Figure 5.

| Depression Level | Frequency/ Hz | Loudness/dB |
|-----------------------|---------------|-------------|
| Minimal/No depression | >184.136 | <58.167 |
| Mild depression | 184.136 | 58.167 |
| Moderate depression | 179.412 | 58.987 |
| Severe depression | 172.316 | 60.216 |

Figure 5. Markers for depression

| | | Depression Level | Number of pauses |
|------------------|-----------------------------|------------------|------------------|
| Depression Level | Pearson | 1.00 | .38 |
| | Correlation Sig. (2-tailed) | | .20 |
| | N | 13 | 13 |
| Number of pauses | Pearson | .38 | 1.00 |
| | Correlation Sig. (2-tailed) | .20 | |
| | N | 13 | 13 |

Figure 6. Correlation between depression level and number of pauses

Figure 6 shows that the bivariate correlation between the depression level and number of pauses was moderate, $r(11) = 0.38$. 14.44% of the variability in the patients' number of pauses can be accounted for by their depression level. The positive coefficient indicates that as depression level increases, number of pauses increases as well. It was also found that people with lower depression scores speak for a longer period of time.

Additionally, the last feature the game detects would be the verbal reaction time in s (GDS) through the reading of one-letter words flashed on the screen. Even though the depression videos could not analyse the verbal reaction time using *Praat* (see Figure 7), this marker was found as an effective indicator of depression from our literature review. The range between 0-4 GDS shows that the level of depression is minimal or void, but a GDS above 4 indicates that the level of depression is relatively high.

B. Construction of Prototype: *A-maze-ing*

A mockup of the game has been made with *Justinmind* and published on *Justinmind Usernote*, a platform for prototype-sharing and evaluation. The mock up was also done on *Powerpoint* (Figures 8 to 11) with the same features as the design on *Justinmind*, except that the prototype has an extra function of recording the player's voice. After that, the voice analysis software called *Praat* will analyse the recorded voice based on pitch, loudness and number of pauses. At a particular junction, the player's reaction time will also be assessed. The results will then be compared to depression markers to determine BDI depression level. Even though we are unable to incorporate *Praat* into the *Powerpoint*, this is how our future app could work.

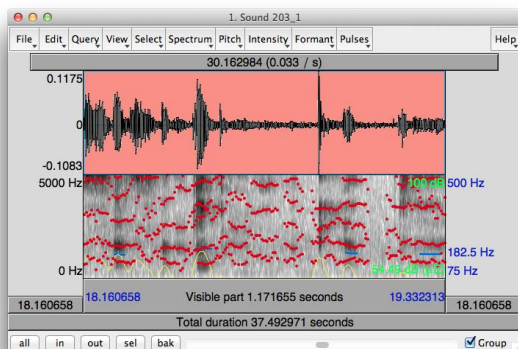


Figure 7. Image showing *Praat* software analysing frequency and loudness of voice but not verbal reaction time.

The Grandfather Passage, a standard 132 word paragraph commonly used in the assessment of communication disorders, was chosen to be included in the game because it had a balance of passage length with breadth of tasks for clinical efficiency, with comprehensive phonotactic coverage to examine speech repertoire [5].

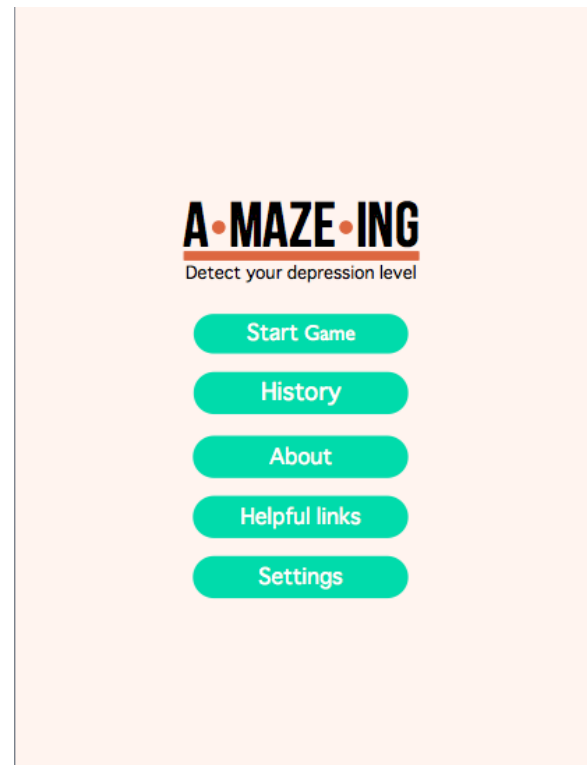


Figure 8. App Interface

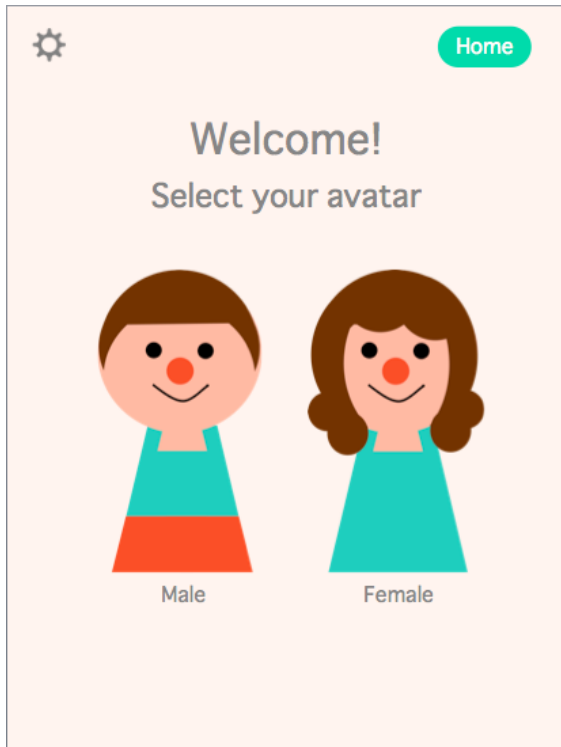


Figure 9. App Interface

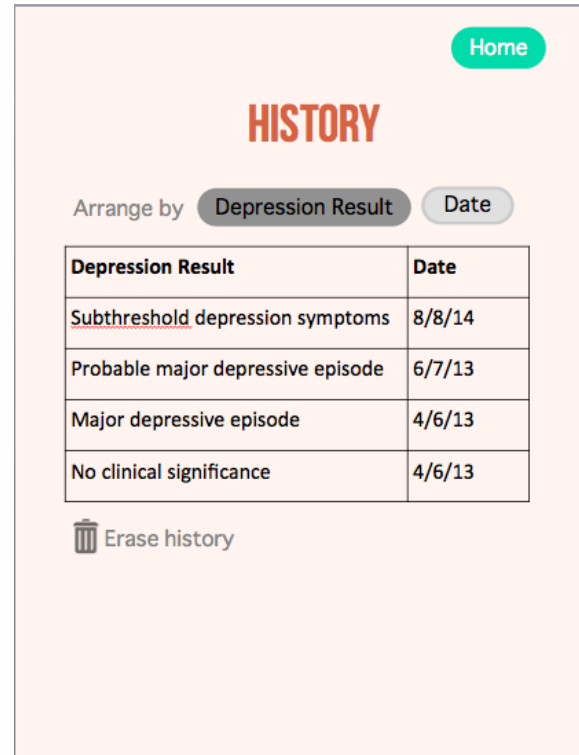


Figure 11. App Interface

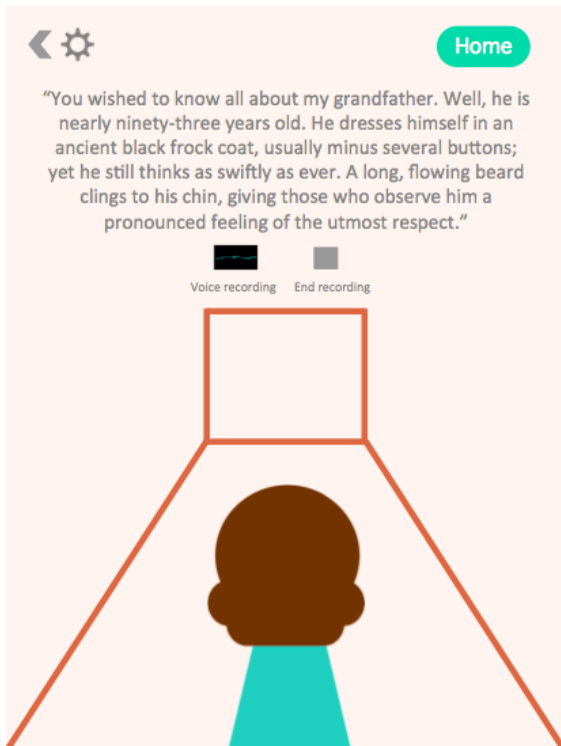


Figure 10. App Interface

IV. DISCUSSION

A. Video Analysis

12 to 13 depression videos from the *AVEC2014* dataset [4] were analysed for various voice features such as loudness, frequency, number of pauses using the software called *Praat*. The statistics prove that, loudness, frequency and number of pauses have moderate correlation to depression, with bivariate correlations of 0.32, -0.31 and 0.38 respectively. The relationship is that the louder an individual's speech, the lower the frequency of the speech and the higher the number of pauses in the speech, the more likely the person is to have depression. From such trends, markers indicating mild depression were derived and to be included in the prototype.

However, the videos were not divided separately based on gender and this could have affected the accuracy of the markers as males generally have lower voices. The patients were mostly adults whereas the target audience is female teenagers. This was due to limited access to depressed teenage patients. Nevertheless, it was assumed that the differences in voice characteristics would be minimal across all genders and age groups and therefore the markers and the application could be applied to the general population.

B. Evaluation of App

Altogether, the prototype has been completed and the user gets to go through the maze while reading The Grandfather Passage. Although we are unable to incorporate *Praat* into the prototype right now, once the app is encoded, *Praat* will work to analyse the speech features while the user is reading the passage, and the scores derived from *Praat* will be based on our markers of depression analysed from the depression videos. Furthermore, as the users continue playing, their verbal reaction time when reading the one-letter words will also contribute to their depression score released at the end of the game. From there, they will know whether they have minimal/no depression, mild depression, moderate depression, or severe depression.

C. Limitations

The app could not be constructed in time due to the complication of encoding which will take up much time. Furthermore, gathering information was crucial before we even started on the app, therefore encoding the app would be under our future work.

V. CONCLUSION

Based on the depression videos, a louder voice, lower frequency of speech, and higher number of pauses are features that one can look out for depression symptoms. The game that we have designed, incorporates all the above information by using females of age 13-19 as our target audience of our game app, and using the above features to help detect for depression in these high-risk youths.

VI. FUTURE DIRECTIONS

Once the encoding of the game is completed, the game will serve as a depression detector for young females of ages 13-19 who are most at risk of depression. This will allow for early detection to ensure that the females receive treatment before the situation worsens. Furthermore, using the survey

results that confirms the relationship between the increasing usage of social media bringing about hints of depression for young females, another app can possibly be programmed to decrease the screen time usage of these frequent internet users by locking their phone after a certain amount of time. This would restrict their number of hours spent on social media and reduce their amount of stress hormones produced when they are on the internet, hence alleviating the likelihood of people getting depression.

VII. ACKNOWLEDGEMENTS

We would like to thank the Institute for Infocomm Research (I²R), A*STAR, and the Science Mentorship Programme, Ministry of Education, Singapore for the opportunity and support of this research project. We would also like to thank the researchers for the contribution and sharing of the *AVEC2014* dataset used in this research work.

VIII. REFERENCES

- [1] O'Keeffe, G. S., & Clarke-Pearson, K. (2011). "The Impact of Social Media on Children, Adolescents, and Families", *Pediatrics: Official Journal of the American Academy of Pediatrics*, 800-804.
- [2] Mok, W.T., Sing, R., Jiang X. T., & See, S.L. (2014). "Investigation of Social Media on Depression", in the proceedings of ISCSLP 2014, 12-14 September, Singapore.
- [3] Beck, A.T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J., (1961). "An inventory for measuring depression", *Archives of General Psychiatry*, 561-571.
- [4] Valstar, M. F. , Schuller, B., Smith, K., Al-maev, T., Eyben, F., Krajewski, J., Cowie, R., & Pantic, M. (2014). AVEC 2014 3d dimensional affect and depression recognition challenge. In *International Conference on ACM Multimedia Audio/Visual Emotion Challenge and Workshop*.
- [5] Gray, W. S. (1936). *Standard oral reading paragraphs: My grandfather*. Bloomington, IL: Public School Publishing Company.