Relationship between Voice Characteristics and Mental Health

Liu Tsz Huen, Isabella

Supervised by Dr. Estella Ma

The University of Hong Kong

Laidlaw Scholars Program 2023

香 港 大 學
THE UNIVERSITY OF HONG KONG

Acknowledgements

This project is a remarkable start to my journey in enhancing the current practices of speech-language pathologists as a budding researcher. This research would not be possible without the generous support of the Laidlaw Scholars Foundation for placing trust in me to incubate this idea in the raw.

I would also like to show my sincere appreciation to Dr. Estella Ma for her guidance and patience. Her insights and suggestions have guided the advancement of the research process.

I am grateful for the contribution of all my participants, whom I have talked to and shared the privilege of learning about their stories. They have devoted their time and effort to actualizing the vision of improving mental healthcare.

I would also like to give a token of thanks to my friends and family for the unwavering support they have provided throughout this journey.

Abstracts

This research explored the use of voice characteristics as an objective tool for diagnosis and tracking of recovery progress for individuals with mental health disorders, and this study particularly focused on the condition of depression. This study explored the use of different speech types with varying emotional valence and intensity to collect voice samples for the differentiation between individuals with or without depression. Despite robust research from Western countries and mainland China, there has yet to be an investigation on the Cantonese-speaking population. Since Cantonese is a tonal language with a very distinct and unique phonological system, it can yield different results compared to previous studies. By classifying the voice tasks into two categories of emotionally arousing and emotionally neutral tasks, the author explored the effectiveness of various tasks in differentiating the difference in voice characteristics among the experimental and control groups. A significant difference between the two groups was found in using an emotionally arousing task with less emotional intensity and positive valence (i.e., reading aloud a passage that elicits positive emotions).

Table of Contents

Acknowledgements 2

Abstracts 3

1. Introduction 5

2. Methods 8

2.1 Participants 8

2.2 Procedures 8

2.3 Data Analysis 10

3. Results 12

4. Discussion 16

5. Conclusion 19

1. Introduction

Mental health problems are pervasive among the population in Hong Kong. The prevalence of common mental disorder in Hong Kong adults was 13.3% according to the Hong Kong Mental Morbidity Survey 2010 – 2013. Mixed anxiety and depressive disorder (6.9%) and depressive episode (2.9%) are one of the most common disorders (Lam et al., 2015).

However, the existing tools for diagnosis, such as the Hamilton Rating Scale for Depression (Hamilton, 1960) and Beck Depression Inventory (Beck et al., 1961), were deemed over-reliance on subjective perception from clinicians or patients themselves. Hence, emerging research investigates a set of biological, physiological, and behavioural markers to aid clinical assessment.

Acoustic features of speech have been well-researched and identified as objective markers of patients with major depressive disorder. Research suggests that anxious and depressive voices are characterized by lower pitch, pitch range, and speaking intensity. (Darby & Hollien, 1977; Hollien, 1980) Particularly with a reduced fundamental frequency (F0) range and a reduced F0 average (Nilsonne & Sundberg, 1985). Patients with mental health disorders particularly depression are generally described as having a monotone which results from psychomotor retardation, i.e., slowing of thoughts and physical movement (Horwitz et al., 2013; Quatieri & Malyska, 2012). Since the larynx is an intricate neuromuscular system, there can be a disturbance in laryngeal tension and control caused by psychomotor retardation, leading to a reduction in F0 variability (Breznitz, 1992; Cannizzaro et al., 2004)

Apart from the widely researched acoustic features, speech characteristics (e.g., linguistic stress and articulatory vowels) have also been investigated in some research (Stasak et al., 2019)

Previous findings suggested a reduced speaking intensity, slower speech, reduced intonation, a lack of linguistic stress, and more articulation errors in depressive speech (Cummins et al., 2015).

Although there has been a number of research investigating the characteristic of voice and speech of individuals with depression, particularly from the Western population (Kuny & Stassen, 1993; Mundt et al., 2007) and China (Wang et al., 2019; Zhao et al., 2022), there has not been any research conducted or database collection in the Cantonese speaking community.

Given that Cantonese is a language with a rich tonal inventory of six lexical tones, it is unknown whether the characteristics of speech samples from Cantonese-speaking individuals can be different from the currently investigated population.

There is little research investigating using different speech types and emotions as classifiers to detect depression, except one research based in Beijing (Jiang et al., 2017). Previous research suggested that tasks that require the recall of autobiographical memories can cause more significant changes in vocal expression compared to reading speech (Alghowinem et al., 2013). Given that different voice tasks, classified into two categories of emotionally neutral and emotionally arousing tasks can be used for voice and speech sample collection, there is missing information on which is a more effective tool for detecting individuals with mental health disorders.

Moreover, individuals with mental health disorders were found to have higher pitch variability (Shin et al., 2021). On the clinical level, understanding the intra-speaking variability of a voice measure of the client's performance across time can reflect the real changes due to treatment or variability of the measure (Ma & Lam, 2015). Since there is a lack of comparison regarding the intra-speaking variability of parameters for individuals with or without mental health disorders,

this will be a piece of important missing information when it comes to designing voice measures specifically for individuals with mental health disorders.

Aim of the research

Due to a lack of a database for voice and speech characteristics in Cantonese-speaking individuals with and without depression in Hong Kong, the topic is considered an uncharted area for the local community. Yet, the prevalence of clinically diagnosed depression in Hong Kong is high, and there is a need for the development of a more objective and effective mental health disorders disorder classification system. This research can help bridge the gap by providing data based on the local community. Moreover, this research can provide details for developing voice measures that effectively detect individuals with mental health disorders by considering the intra-speaking variability of different measures elicited from the voice tasks.

Thus, the objective of this research is (1) to understand the difference in quantitative measurement of voice characteristics in Cantonese-speaking individuals with or without mental health disorders and (2) to explore the use of emotionally arousing or emotionally neutral tasks in voice sample collection. The research questions are (1) What is the difference in quantitative measurement of speech and voice characteristics in Cantonese-speaking individuals with or without mental health disorders? (2) Are the difference in voice characteristics found in Q1 consistent across emotionally neutral and emotionally arousing tasks?

2. Methods

2.1 Participants

The research was separated into two phases – Phase One and Phase Two.

For phase one, a total of 9 participants have been recruited. Participants with a background in psychology, particularly clinical psychology, neuroscience, counselling, mental health, and cognitive science, were recruited in this research.

For phase two, 31 participants (26 Females; 5 Males) were involved in the data collection procedure. Their age range from 19 to 49 (Mean = 24.08, SD = 6.24). Due to the lack of male participants with mental health disorders, only the voice samples of the female participants were included in the data analysis. Among all the participants, 16 formed the control group (i.e., without any mental health disorders), and 8 formed the experimental group (i.e., diagnosed with mental health disorders). They are mostly recruited through friend referrals, promotion on social media (Instagram of Voice Research Laboratory), and sign-ups from mass email promotion. The experimental group indicated that they had received a diagnosis of depression on the sign-up form. The researcher also noted whether they did or did not intake any anti-depressants currently. It is noted that the method of convenience sampling is used in this study.

2.2 Procedures

For phase one, an online survey was distributed and filled in by the participants at any time and location of their convenience during a given period. They were asked to rate ten types of voice tasks on an 11-point rating scale on whether the voice task is emotionally arousing

or emotionally neutral. All of the data collected from the questionnaire were used for analysis purposes.

After receiving the results from part A, the research selected voice tasks categorized as emotionally neutral and emotionally arousing for part B data collection.

For phase two, both the experimental and control groups received the same set of instructions from the researcher throughout the data collection process. Both groups were engaged in a 30-minute voice sample collection. The participants were each asked to perform in total four tasks. Firstly, they were asked to produce the maximum phonation of the /a/ sound for three times. Then, they were instructed to count from 1 to 10 for three times. After that, they were given two passages (conveying negative and positive emotions) to read. Lastly, they were given time to answer two questions regarding their personal experience. An AI-generated audio was used to instruct the participants for the last interview task to avoid any bias caused by the variations in tone or expression used by the researcher while asking the interview questions.

Data collection setting

The data collection took place in a sound booth at Meng Wah Complex at the University of Hong Kong. The noise level was monitored below 35 dBA, measured with a Quest Electronics Permissible sound level meter (model 215). All the voice samples were recorded to a free, open-source audio software Audacity, using a professional grade, dynamic microphone (Shure, Beta 87) at a 15-cm mouth-to-microphone distance. This is to prevent any burst of air caused by the production of plosive sounds during the voice

collection, which may affect the acoustic analysis of the voice samples. The participants were encouraged to speak at a comfortable loudness and pitch.

Selection of voice task stimuli

Among the four tasks that were selected to collect the voice sample, two sets of materials were selected for the third task on passage reading. One of the passages conveyed a negative emotion and was chosen from an authentic story recorded in the project - 100 stories by the headwind team from the Department of Psychiatry at the University of Hong Kong. The stories are available to the public via their official website. This piece of material was converted from written to oral language, with its content trimmed down to match the length of another passage. As for the passage with negative content, it is selected from research studying the impact of emotional intensity on the comprehension and memorization of texts. The passage is a joyous description of a proud sister included in the appendix of the research as mentioned above and transcribed into Cantonese version.

The two research questions resembled those being asked in research suggesting using voice acoustic to measure the severity and treatment response of individuals with depression using the interactive voice response technology. This kind of question resembles the questions being asked in HAMD, a questionnaire commonly used to diagnose patients with depression. Two questions were selected, one question eliciting positive emotions and one eliciting negative emotion.

2.3 **Data Analysis**

After all the voice samples were collected, they were extracted and processed using Praat, a freeware program for the analysis of acoustic speech signals. Different acoustic

parameters, including pitch and intensity, were extracted. Since the emotionally neutral tasks (tasks 1 and 2) were repeated three times, a coefficient of variance (CoV) was calculated using the mean fundamental frequency (F0) and intensity (dB) of the three trials. Information about the spectrograms of the voice samples, including jitter, shimmer, and mean harmonic-to-noise ratio, was also extracted from Praat's Python Library (Parselmouth).

As for the speech rate, it was first transcribed to calculate the word count to find out the syllables per second (wpm) as a unit for measurement. Two AI models were utilized to recognize the transcripts of the voice samples (tasks 3 and 4) on Google Colab with Python for cross-validating to ensure the robustness of the process.

After all the parameters have been processed and extracted, they are inputted into the Statistical Product and Service Solutions (SPSS) for further analysis. The normality of the data was first checked to make sure there was a normal distribution. Then, independent-sample t-tests were run to compare the different acoustic parameters of individuals with or without depression. By comparing the mean of the two groups, we can test whether the null hypothesis should be rejected or not. The mean, standard deviation (SD), p-value, and degree of freedom were reported to indicate whether there is a significant difference in the acoustic parameters of the two groups. Then, the effect size represented by Cohen's d was also reported to indicate the importance of the effect found.

For the emotionally neutral tasks, the mean pitch (F0), mean intensity (dB), and Coefficient of Variation (CoV) were reported since the two tasks were repeated for 3 trials. As for the emotionally arousing tasks in the format of connected speech, the parameters: mean pitch (F0), mean intensity (dB), and speech rate in syllables per second were reported.

3. Results

Table I. Rating of voice tasks commonly used by Speech therapist

	Tasks	Type (N: neutral; A: Arousing)	Rating (median)
1	Counting from 1 to 10	N	0
2	Reading aloud short passage (The North Wind and the Sun)	N	1
3	Diadochokinetic rate	N	1
4	Reading aloud sentences	N	2
5	Prolonged /a/	N	3
6	Words Reading	N	4
7	Answering personally-related questions (Interview)	A	5
8	Reading aloud short passages (Conveying negative and positive emotions)	A	5
9	Picture Description	A	6
10	Video Viewing	A	7

For phase one, in total 9 responses were collected through the questionnaire. Out of all the voice tasks, the respondents rated "Reading aloud short passage (The North Wind and the Sun)," "Measurement of Diadochokinetic rate," and "Counting from 1 to 10" to be the most emotionally neutral tasks. While "picture description" and "video viewing" were the most emotionally arousing tasks.

Due to factors in execution, including copyright of the materials and the consideration of the utility of these tasks employed by speech therapists, four tasks: "Interview with the client," "Prolonged /a/", "Passage reading (emotionally arousing texts)" and "Counting from 1 to 10" were included in the studies. Nonetheless, the researcher is open to further exploring different voice tasks not included in this research.

Table II. Difference in acoustic parameters between the experimental and control group for emotionally neutral tasks

		Experimental	Control	t(22)	p-value
1.	Mean Pitch (F0) of Task 1 - Prolonged /a/				
	Mean	231.42	224.68	-0.47	0.64
	SD	28.08	35.20		
2.	Mean Pitch (F0) of Task 2 – Counting from 1 to 10				
	Mean	223.80	215.05	- 0.92	0.37
	SD	29.52	17.34		
3.	Coefficient of Variation (CoV) of Task 1 – Prolonged /a/				
	Mean	0.20	0.03	0.50	0.62
	SD	0.02	0.04		
4.	Coefficient of Variation (CoV) of Task 2 – Counting from 1 to 10				
	Mean	0.03	0.03	0.14	0.89
	SD	0.01	0.02		
5.	Mean Intensity of Task 1 – Prolonged /a/				
	Mean	63.47	62.20	- 0.41	0.69
	SD	5.13	7.97		
6.	Mean Intensity of Task 2 – Counting from 1 to 10				
	Mean	50.88	49.05	- 1.21	0.24
	SD	4.77	2.71		

Regarding the result of phase two, there is no significant difference with $p > 0.05$ in acoustic parameters between individuals with or without depression using emotionally neutral tasks (i.e., prolonged /a/ and counting from 1 to 10) as the means of voice samples collection.

Despite a non-significant difference was found, it was generally observed that the mean pitch and mean intensity of the experimental group in performing the two emotionally neutral tasks were higher than that of the control group.

In addition, the Coefficient of Variation (CoV) of Task 2 for the experimental group was lower than that of the control group. Since lower CoV suggested a larger variation of the data among the 3 trials, this can be taken into consideration for further research. It is noted that a small sample size of this research can also affect the validity of the resulting data.

Table III. Difference in acoustic parameters between experimental and control group for emotionally arousing tasks that elicited negative emotions

Negative Emotions Elicited					
		Experimental	Control	t(22)	p-value
1.	Mean Pitch (F0) of Task 3 – Passage Reading				
	Mean	218.43	220.47	0.20	0.84
	SD	23.10	23.43		
2.	Mean Pitch (F0) of Task 4 – Interview Question				
	Mean	224.45	217.08	- 0.56	0.58
	SD	21.05	34.26		
3.	Mean Intensity of Task 3 – Passage Reading				
	Mean	57.09	54.76	- 1.91	0.07
	SD	3.83	2.19		
4.	Mean Intensity of Task 4 – Interview Question				
	Mean	53.85	53.26	- 0.38	0.71
	SD	5.28	2.51		
5.	Speech Rate of Task 3 – Passage Reading				
	Mean	3.68	3.78	0.53	- 0.61
	SD	0.54	0.38		
6.	Speech Rate of Task 4 – Interview Question				
	Mean	3.60	3.66	0.29	0.77
	SD	0.37	0.52		

The acoustic parameters between individuals with or without depression using emotionally arousing tasks that elicited negative emotions were all found to be insignificant, with $p > 0.05$.

Despite insignificant differences, it is noted that, in general, the voice samples of the experimental group have higher mean pitch and mean intensity than the control group, while the speech rate of the two tasks was generally lower than that of the control group.

The tendency of difference in mean pitch and mean intensity were aligned with that of the acoustic parameters data collected using emotionally neutral tasks, which deserved further investigation.

Table IV. Difference in acoustic parameters between experimental and control group for emotionally arousing tasks that elicited positive emotions

Positive Emotions Elicited						
		Experimental	Control	t(22)	p-value	Effect Size (d)
1.	Mean Pitch (F0) of Task 3 – Passage Reading					
	Mean	224.70	222.37	- 0.22	0.82	/
	SD	29.99	20.54			
2.	Mean Pitch (F0) of Task 4 – Interview Question					
	Mean	225.88	219.57	- 0.48	0.63	/
	SD	27.35	31.42			
3.	Mean Intensity of Task 3 – Passage Reading					
	Mean	57.97	55.03	- 2.33	0.03*	0.96 (large)
	SD	3.42	2.64			
4.	Mean Intensity of Task 4 – Interview Question					
	Mean	54.53	53.18	- 0.90	0.38	/
	SD	4.85	2.59			
5.	Speech Rate of Task 3 – Passage Reading					
	Mean	3.19	4.13	2.53	0.02*	1.23 (large)
	SD	0.41	1.00			
6.	Speech Rate of Task 4 – Interview Question					
	Mean	3.60	3.86	1.06	0.30	/
	SD	0.54	0.57			

Regarding the acoustic parameters of individuals with and without depression recorded using emotionally arousing tasks that elicited positive emotions, two parameters namely the mean intensity of passage reading showed significant difference with $p < 0.05$ and a large effect size ($d = 0.96$), and the speech rate of passage reading also showed a significant difference with $p < 0.05$ and a large effect size ($d = 1.23$).

While for the voice samples recorded using another emotionally arousing task (interview question) that elicited negative emotions do not show any significant difference in any acoustic parameters. It was again generally observed that the mean pitch and intensity of the experimental group were higher than that of the control group while the speech rate were lower than that of the control group.

4. Discussion

The acoustic parameters generally demonstrated an aligned pattern across the four voice tasks. The experimental group's mean pitch and mean intensity were higher than the control groups. As for the speech rate, the experimental group showed a slower rate of producing these outputs than the control group. This overall pattern painted the basis of the discussion.

It was hypothesized that the emotionally neutral tasks are less likely to create any emotional changes as they carry less emotional intensity than the emotionally arousing tasks. Hence, it is reasonable that no significant differences were found among the acoustic parameters. From previous studies (Kuny & Stassen, 1993; Zhao et al., 2022), the difference in acoustic parameters was found despite an emotionally neutral task such as reading an emotionally neutral passage or counting aloud from 1 – 30. However, it is noted that the participants from these studies were recently hospitalized and in the process of treatment using anti-depressants. While the participants in this research were in a stage of recovery without being in an acute stage, most of them were also on an anti-depressant prescription. Given that previous studies proposed that participants showed improvement in their depression symptoms with the acoustic measures showing a similar trend of diminishing severity over time (Jiang et al., 2017; Kuny & Stassen, 1993), the non-significant can be explained by the intake of anti-depressant that will cause a reduced disturbance in laryngeal tension and control caused by psychomotor retardation (Mundt et al., 2007; Quatieri & Malyska, 2012), which is the leading cause of acoustic difference between individuals with and without depression when emotions were not induced in the voice tasks.

We can then compare the two voice tasks among the emotionally arousing tasks: passage reading and interview questions. It was observed that differences in acoustic parameters

were only found to be significant in passage reading rather than the interview questions. The non-significance can be explained by the nature of the tasks used to elicit the voice samples. For the interview questions, a recall of autobiographical memories was required when the participants were asked to illustrate personal events (Alghowinem et al., 2013). This task tended to carry more emotional intensity than tasks that do not require a recall of autobiographical memories (i.e., passage reading). Hence, it is more likely to trigger more significant changes in the vocal expression of the participants than reading speech. Under this circumstance, the experimental and control groups showed similar voice characteristics when illustrating personal events. This led to non-significance across acoustic parameters from voice samples recorded using interview questions.

Regarding the passage reading task, there is a significant difference in acoustic parameters across participants with and without depression for the passage that elicited positive emotions. While for the passage that elicited negative emotions, no significant differences were found. This is an important finding given that the type of emotions (i.e., emotional valence) elicited from the voice tasks can also impact the acoustic characteristics of the individuals. Out of the parameters, the difference in intensity and speech rate was observable among the two groups. Both parameters were greatly related to the reduced control and accumulated tension in the laryngeal muscle that is characterized in individuals with depression due to psychomotor retardation. The dysfunction of GABA neurotransmitters (Croarkin et al., 2011; Ellgring & Scherer, 1996) in individuals with depression will increase muscle tension. Hence, a higher intensity and lower speech rate displayed by the experimental group can be explained by the weakness in motor coordination and increased laryngeal muscle tension due to depression. Moreover, the use of a positive passage can cause such difference in voice characteristics rather than a

negative passage because it is harder for the experimental group to mimic an energetic and loving voice for reading aloud passages with positive valence that are not personally relevant to them as it arouses less emotional intensity (Megalakaki et al., 2019).

In conclusion, the above factors explained the difference in the acoustic parameters of voice samples elicited by different voice tasks that carry different emotional intensity and valence.

5. Conclusion

This research provided an initial investigation regarding using different speech types and emotions as classifiers for diagnosing and treating mental health disorders using voice characteristics as an objective tool. Indeed, reading aloud passages that convey positive and negative emotions can be an effective tool for identifying individuals with or without depression, given that a significant difference can be yielded from reading a passage that elicits positive rather than negative.

It is noted that this research acts as a preliminary investigation into the concept of voice characteristics and mental health, which is a topic worth our deep concerns and attention, especially after the brutal hit by the pandemic. There are many research potentials and investigations regarding this area of exploration. Based on this existing research, since a convenient sampling method was used, a more precise selection of participants can be included in future research. To illustrate, the diagnosis of individuals with depression can be confirmed by a diagnosis from the clinical psychologists or the collection of case history. It is also worth exploring the difference in voice characteristics among individuals with depression regarding the type of medication they took and the status of their recovery. Keeping the background consistent across participants will help eliminate the confounding factors that can impact the research data. Moreover, it is worth spending more effort in recruiting male participants to explore the factor of gender on the relationship between voice characteristics and mental health..

With the notion of contributing to the advancement of healthcare and well-being of individuals with mental health disorders, further research can be focused on exploring the effectiveness and nature of other voice tasks that are not employed in this study (e.g.,

picture description and video commenting tasks). The effectiveness of different classifiers can be compared and contrasted.

In addition, it is also worth exploring the use of acoustic parameters to monitor the treatment progress of individuals with mental health disorders, such as comparing medication effects across different baselines to find the most suitable medication for different individuals. Mental health disorders besides depression, such as anxiety and schizophrenia, can be different directions for future exploration.

Lastly, it is believed that with the initial voice database accumulation and the development of AI models for data analysis, this research topic can be further advanced with the use of different technologies, such as machine learning and the development of software for more convenient usage of voice characteristics as a mental health diagnostic or severity measurement tool.

References

Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Breakspear, M., & Parker, G. (2013). *Detecting Depression: A Comparison between Spontaneous and Read Speech*. https://doi.org/10.1109/ICASSP.2013.6639130

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Arch Gen Psychiatry*, *4*, 561-571. https://doi.org/10.1001/archpsyc.1961.01710120031004

Breznitz, Z. (1992). Verbal Indicators of Depression. *The Journal of General Psychology*, *119*(4), 351-363. https://doi.org/10.1080/00221309.1992.9921178

Cannizzaro, M., Harel, B., Reilly, N., Chappell, P., & Snyder, P. J. (2004). Voice acoustical measurement of the severity of major depression. *Brain Cogn*, *56*(1), 30-35. https://doi.org/10.1016/j.bandc.2004.05.003

Croarkin, P. E., Levinson, A. J., & Daskalakis, Z. J. (2011). Evidence for GABAergic inhibitory deficits in major depressive disorder. *Neurosci Biobehav Rev*, *35*(3), 818-825. https://doi.org/10.1016/j.neubiorev.2010.10.002

Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication*, *71*, 10-49. https://doi.org/https://doi.org/10.1016/j.specom.2015.03.004

Darby, J. K., & Hollien, H. (1977). Vocal and speech patterns of depressive patients. *Folia Phoniatr (Basel)*, *29*(4), 279-291. https://doi.org/10.1159/000264098

Ellgring, H., & Scherer, K. R. (1996). Vocal indicators of mood change in depression. *Journal of Nonverbal Behavior*, *20*(2), 83-110. https://doi.org/10.1007/BF02253071

Hollien, H. (1980). Vocal indicators of psychological stress. *Ann N Y Acad Sci*, *347*, 47-72. https://doi.org/10.1111/j.1749-6632.1980.tb21255.x

Horwitz, R., Quatieri, T., Helfer, B., Yu, B., Williamson, J., & Mundt, J. (2013). *On the relative importance of vocal source, system, and prosody in human depression*. https://doi.org/10.1109/BSN.2013.6575522

Jiang, H., Hu, B., Liu, Z., Yan, L., Wang, T., Liu, F., Kang, H., & Li, X. (2017). Investigation of different speech types and emotions for detecting depression using different classifiers. *Speech Communication*, *90*, 39-46. https://doi.org/https://doi.org/10.1016/j.specom.2017.04.001

Kuny, S., & Stassen, H. H. (1993). Speaking behavior and voice sound characteristics in depressive patients during recovery. *Journal of Psychiatric Research*, *27*(3), 289-307. https://doi.org/https://doi.org/10.1016/0022-3956(93)90040-9

Lam, L. C., Wong, C. S., Wang, M. J., Chan, W. C., Chen, E. Y., Ng, R. M., Hung, S. F., Cheung, E. F., Sham, P. C., Chiu, H. F., Lam, M., Chang, W. C., Lee, E. H., Chiang, T. P., Lau, J. T., van Os, J., Lewis, G., & Bebbington, P. (2015). Prevalence, psychosocial correlates and service utilization of depressive and anxiety disorders in Hong Kong: the Hong Kong Mental Morbidity Survey (HKMMS). *Soc Psychiatry Psychiatr Epidemiol*, *50*(9), 1379-1388. https://doi.org/10.1007/s00127-015-1014-5

Ma, E. P., & Lam, N. L. (2015). Speech task effects on acoustic measure of fundamental frequency in Cantonese-speaking children. *Int J Pediatr Otorhinolaryngol*, *79*(12), 2260-2264. https://doi.org/10.1016/j.ijporl.2015.10.018

Megalakaki, O., Ballenghein, U., & Baccino, T. (2019). Effects of Valence and Emotional Intensity on the Comprehension and Memorization of Texts [Original Research]. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00179

Mundt, J. C., Snyder, P. J., Cannizzaro, M. S., Chappie, K., & Geralts, D. S. (2007). Voice acoustic measures of depression severity and treatment response collected via

interactive voice response (IVR) technology. *J Neurolinguistics*, *20*(1), 50-64. https://doi.org/10.1016/j.jneuroling.2006.04.001

Nilsonne, Å., & Sundberg, J. (1985). Differences in Ability of Musicians and Nonmusicians to Judge Emotional State from the Fundamental Frequency of Voice Samples. *Music Perception*, *2*(4), 507-516. https://doi.org/10.2307/40285316

Quatieri, T., & Malyska, N. (2012). Vocal-source biomarkers for depression: A link to psychomotor activity. *Proceedings of Interspeech*, *2*, 1059-1062.

Shin, D., Cho, W. I., Park, C. H. K., Rhee, S. J., Kim, M. J., Lee, H., Kim, N. S., & Ahn, Y. M. (2021). Detection of Minor and Major Depression through Voice as a Biomarker Using Machine Learning. *J Clin Med*, *10*(14). https://doi.org/10.3390/jcm10143046

Stasak, B., Epps, J., & Goecke, R. (2019). An investigation of linguistic stress and articulatory vowel characteristics for automatic depression classification. *Computer Speech & Language*, *53*, 140-155. https://doi.org/https://doi.org/10.1016/j.csl.2018.08.001

Wang, J., Zhang, L., Liu, T., Pan, W., Hu, B., & Zhu, T. (2019). Acoustic differences between healthy and depressed people: a cross-situation study. *BMC Psychiatry*, *19*(1), 300. https://doi.org/10.1186/s12888-019-2300-7

Zhao, Q., Fan, H. Z., Li, Y. L., Liu, L., Wu, Y. X., Zhao, Y. L., Tian, Z. X., Wang, Z. R., Tan, Y. L., & Tan, S. P. (2022). Vocal Acoustic Features as Potential Biomarkers for Identifying/Diagnosing Depression: A Cross-Sectional Study. *Front Psychiatry*, *13*, 815678. https://doi.org/10.3389/fpsyt.2022.815678