An Attempt to Estimate Depressive Status from Voice

Yasuhiro Omiya[1,2(✉)], Takeshi Takano[1,3], Tomotaka Uraguchi[1], Mitsuteru Nakamura[2], Masakazu Higuchi[2], Shuji Shinohara[3], Shunji Mitsuyoshi[3], Mirai So[4], and Shinichi Tokuno[2]

[1] PST Inc., Industry & Trade Center Building 905, 2 Yamashita-cho, Naka-ku, Yokohama, Kanagawa, Japan
{omiya, takano, uraguchi}@medical-pst.com
[2] Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
{m-nakamura, higuchi, tokuno}@m.u-tokyo.ac.jp
[3] Graduate School of Engineering, The University of Tokyo, Tokyo, Japan
{shinohara, mitsuyoshi}@bioeng.t.u-tokyo.ac.jp
[4] Ginza Taimei Clinic, Tokyo, Japan
mirai.so@keio.jp

Abstract. In the whole world especially developed countries, increasing mental health disorders is a serious problem. As a countermeasure, the main objective of this paper is an attempt to estimate depressive status from voice. In this study, we gathered patients with major depressive disorders in the hospital's consulting room. Several questionnaires including "the Hamilton Depression Rating Scale" (HAM-D) were administered to evaluate the patients' depressed state. Voices corresponding to three long vowels were recorded from the subjects. Next, the acoustic feature quantity was calculated based on the voice. We developed the HAM-D score estimation algorithm from the voice using one of three types of long vowel audio content. As a result, there was a correlation between the "Actual HAM-D Score" and the "Estimated HAM-D Score". We found that the algorithm is effective in estimating depression state and can be used for estimating the disease state based on voice.

Keywords: Vocal analysis · Depressive status estimation · The Hamilton Depression Rating Scale (HAM-D)

1 Introduction

In the whole world especially developed countries, increasing mental health dis-orders is a serious problem, and thus various screening techniques and countermeasures have been studied. For diagnostic support, medical interviews by specialists (e.g., using "the Hamilton Rating Scale for Depression" [HAM-D] [1]), self-report type psychometric tests (e.g., "the Patient Health Questionnaire" [PHQ-9] [2], and "the Beck Depression Inventory" [BDI] [3, 4] are used to screening depressive status of patients with mental health disorders. However, medical interviews by specialists are limited by the number of patients that can be evaluated, and though self-report type psychometric tests are useful in

P. Cipresso et al. (Eds.): MindCare 2019, LNICST 288, pp. 168–175, 2019.
https://doi.org/10.1007/978-3-030-25872-6_13

determining mental health status at their early stages and in complementing diagnoses, there are issues of reporting biases. Regarding evaluations with biomarkers such as saliva [5, 6] and blood [7, 8], they are invasive, and the tests reagent may be required or analysis may take time, and the tests are expensive. Therefore, those methods are not appropriate as easy or simple solutions. In contrast, voice-based evaluation methods have several advantages; for example, they provide diagnostic support to doctors, are almost non-invasive, no needs special equipment, and there can be used remotely and easily. It is thought that depression is triggered by psychological stress that causes the brain to lose its balance with the stress. We conducted studies to estimate stress condition in patients and support the diagnosis of disease using voice [9, 10]. Research on the relationship analyzed the pose from question to the answer and analyzed the relationship with mental health disorders, as well as the vocal fundamental frequency (F0) [11]. However, in the analysis of conversation, it is necessary to have a talk partner, the test cannot be performed alone, and it is time-consuming.

Because patients can provide false information in interviews and questionnaires, objective indicators are effective for diagnostic support. If the depression state of a patient can be estimated from the voice obtained during examination, the burden on the examiner can be reduced and possibly support the diagnosis. In this paper, our aim was to estimate the depression state of patients from their voicing of long vowels, which do not depend on the native language.

2 Materials and Methods

2.1 Experiment

We recorded the voice of patients with major depressive disorder (MDD) in a hospital's consulting room. In addition, HAM-D was conducted to screen for depressed mood. Further, patients were excluded if they had been diagnosed with serious physical disorders or organic brain disease diagnosed by a psychiatrist using The Mini-International Neuropsychiatric Interview [12]. Subjects vocalized three types of the long vowels, such as /Ah/, /Eh/, and /Uh/ approximately three seconds. The voices recorded as 24bit/96 kHz pcm audio files, using the Portable Recorder "R-26" (Roland, Japan), and a pin microphone "ME-52 W" (OLYMPUS, Japan). As for the utterance content, long vowels were selected because they do not depend on the native language. After the time of second visit, we collected the subjects voice and conducted HAM-D at each visit to observe the progress of the patient's recovery. Since the patients were undergoing treatment by a doctor, in many cases, the symptoms improved. As some patients did not revisit the hospital when their condition improved, their voice could not be recorded.

The data were collected from 28 subjects and the total number of data were 42 (Table 1).

This study was approved by the University of Tokyo, Clinical Research Review Board and informed consent was obtained from all the subjects.

Table 1. Collected subjects information.

Gender	Number of recordings			Age		Number of data	HAM-D score	
	Once	Twice	3 times	Mean ± S.D.	t-test		Mean ± S.D.	t-test
Male	5	5	1	32.5 ± 5.9	n.s.	20	23.3 ± 11.1	n.s.
Female	10	7	0	32.2 ± 8.6		22	25.0 ± 9.3	

2.2 Hamilton Depression Rating Scale (HAM-D)

"The Hamilton Rating Scale for Depression" (HRSD), also called "the Hamilton Depression Rating Scale" (HDRS), abbreviated as HAM-D, is a multi-item questionnaire used to indicate signs of depression, and as a tool to evaluate recovery. HAM-D is used to conduct clinical surveys, and its examination by health care workers usually takes about 15 to 20 min. HAM-D has various items such as "HAM-D-17", "HAM-D-21", "HAM-D-24", and so on.

2.3 Analysis of Data

We used the openSMILE software [13], 6,552 acoustic features were calculated from the collected subjects voice.

The acoustic features were calculated as follows:

(A) Physical feature was calculated on the basis of a frame unit or 56 similar feature types.
(B) Three processing contents (moving average and time change) in the time direction concerning (A)
(C) After the process (B) is performed on (A), statistical quantities of 39 types

Then the feature selection was done by implementing "CfsSubsetEval" and "BestFirst" settings in Weka software [14]. "The Sequential Minimal Optimization" (SMO) algorithm and "the Support Vector Machines" (SVM) for regression algorithm were implemented [15] in Weka software. We used 10-fold cross-validation (10FCV) in order to develop the HAM-D score estimation algorithm, generated from voice associated with pronunciation of each long vowel.

In the analysis, the correlation coefficient between the "Actual HAM-D Score" and the "Estimated HAM-D Score", according to the generated algorithm, was evaluated.

3 Results and Discussion

Figures 1, 2 and 3 are scatter plots for the "Estimated HAM-D Score" versus the "Actual HAM-D Score" based on the acoustic features of /Ah/, /Eh/, and /Uh/.

The performance of the algorithm with 10FCV relative to the "Estimated HAM-D Score" using the acoustic features of /Ah/, /Eh/, and /Uh/, and the correlation coefficients with the "Actual HAM-D Score" were found to be 0.68, 0.83, and 0.68, respectively.

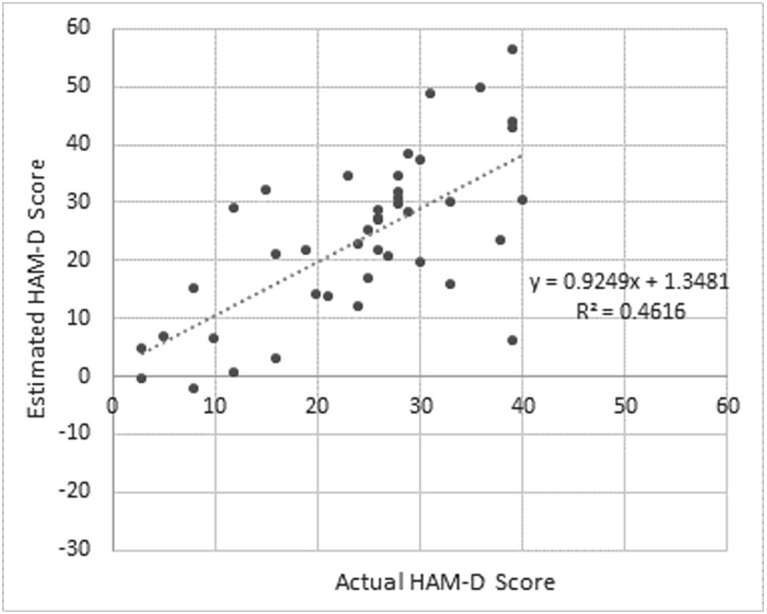

Fig. 1. Scatter plot for the "Estimated HAM-D Score" versus the "Actual HAM-D Score" based on the acoustic features of /Ah/.

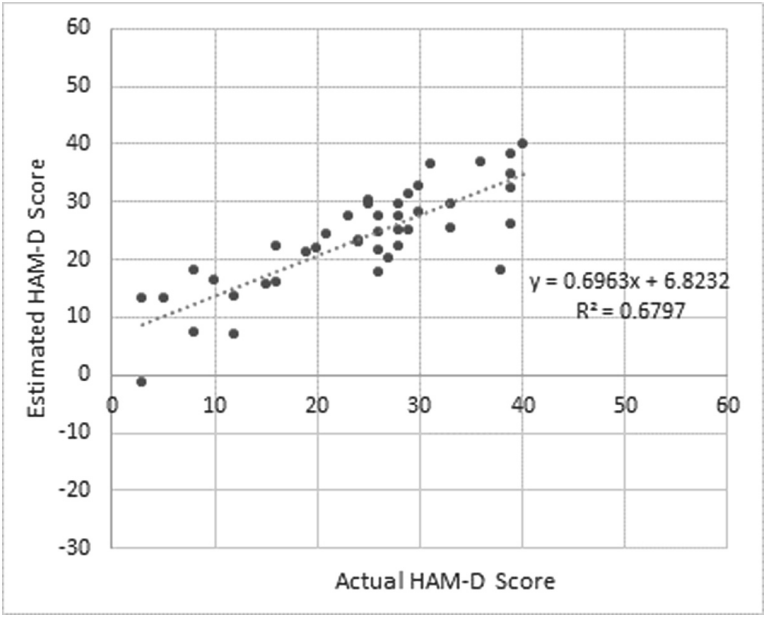

Fig. 2. Scatter plot for the "Estimated HAM-D Score" versus the "Actual HAM-D Score" based on the acoustic features of /Eh/.

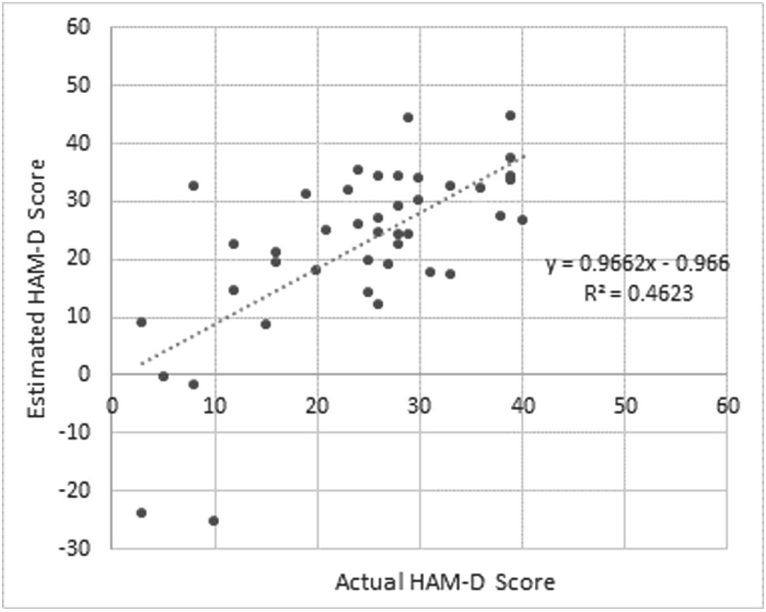

Fig. 3. Scatter plot the "Estimated HAM-D Score" versus the "Actual HAM-D Score" based on the acoustic features of /Uh/.

The correlation coefficient between the "Actual HAM-D Score" and the "Estimated HAM-D Score", using acoustic features of /Eh/, was extremely high at about 0.83. Even when /Ah/and /Uh/results were included, the correlation coefficient was 0.65 or more, and the results suggested that the "Estimated HAM-D Score" based on the acoustic features of the long vowels, /Ah/, /Eh/, and /Uh/, is effective in estimating depression state and can be used for estimating the disease state based on voice.

We then evaluated the algorithm performance to check if the subjects' voices can serve as a parameter to identify those with depressive mental health conditions, as indicated by a HAM-D score of more than 17, which is the cutoff point of "indicates moderate depression" discussed by Zimmerman et al. [16] The receiver operating characteristic (ROC) curve is shown in Figs. 4, 5 and 6 with the vertical axis as sensitivity and the horizontal axis as 1-specificity.

For the "Estimated HAM-D Score" using the acoustic features of /Ah/, /Eh/, and /Uh/, the calculated area under the ROC curve (AUC) values were 0.84, 0.98, and 0.87. This result indicates that the developed algorithm functions correctly in estimating the severity of HAM-D score using the acoustic features of long vowels, and it suggested that it may be effective to estimate mental health status based on voice.

In this study, we used long vowels, which do not depend on the language. Moreover, since our method does not measure the pause duration in conversations, as done in the conventional method, it can easily be used by oneself. However, we acquired voices in a specific recording environment. Therefore, it is necessary to evaluate our method in other recording environments.

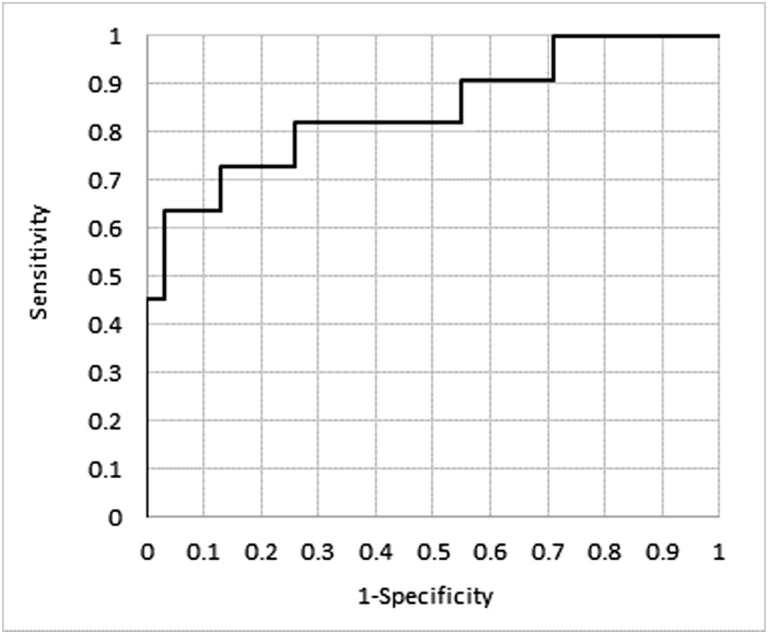

Fig. 4. ROC curve of the "Estimated HAM-D Score" using the acoustic features of /Ah/.

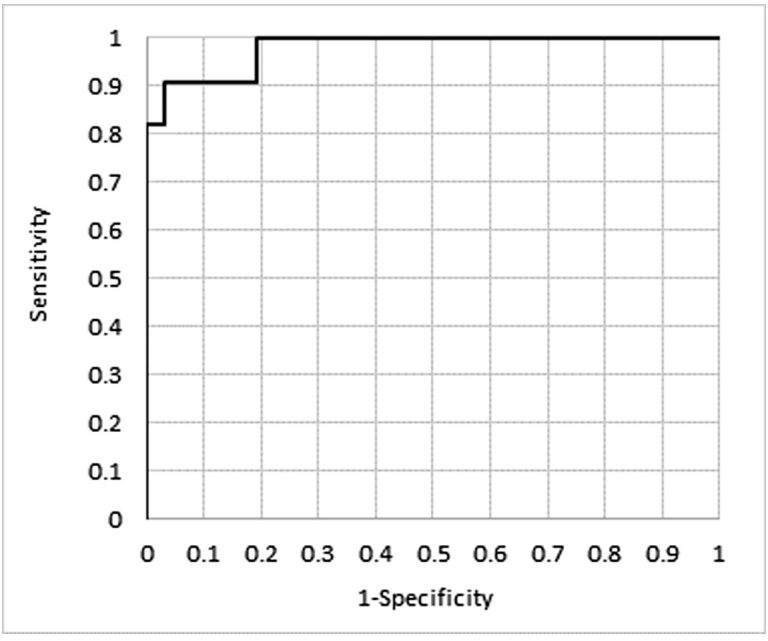

Fig. 5. ROC curve of the "Estimated HAM-D Score" using the acoustic features of /Eh/.

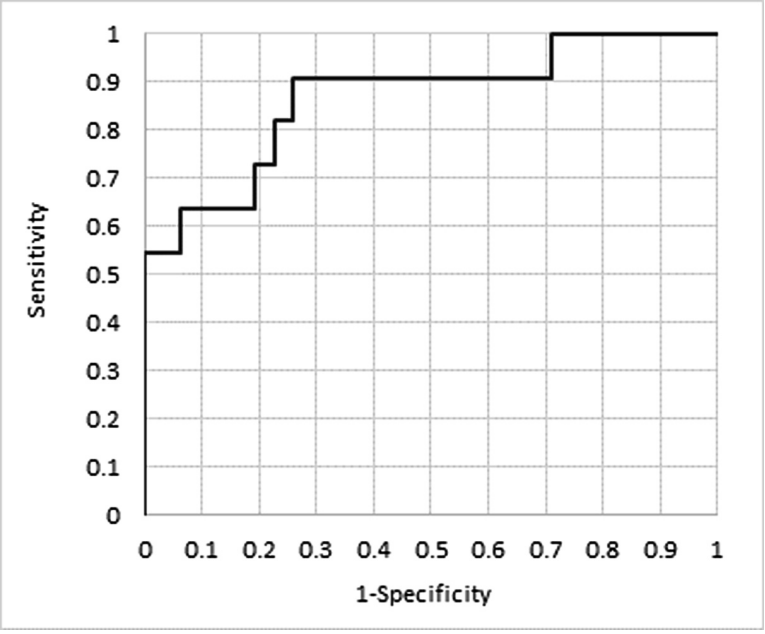

Fig. 6. ROC curve of the "Estimated HAM-D Score" using the acoustic features of /Uh/.

4 Conclusion

In this research, we developed an algorithm to estimate the HAM-D score estimation algorithm from the voice using acoustic features of one of three types of the long vowel such as /Ah/, /Eh/, and /Uh/. Then, we conducted an experiment to compare the HAM-D scores generated from the developed algorithm. The results indicated that the developed algorithm functions correctly in estimating the severity of HAM-D score that indicates depressive status, and can be used for estimating depressive mental health conditions based on voice. Since voice can be easily acquired using devices such as smartphones and personal computers, it is possible to monitor the mental state at home using our algorithm, which could then lead to a doctor's diagnostic support.

Future studies to evaluate the algorithm for other depression-related diseases, such as bipolar with atypical features should be conducted to improve the accuracy.

References

1. Hamilton, M.: Rating depressive patients. J. Clin. Psychiatry **41**, 21–24 (1980)
2. Kroenke, K., Spitzer, R.L., Williams, J.B.: The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. **2001**(16), 606–613 (2001)
3. Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., Erbaugh, J.: An inventory for measuring depression. Arch. Gen. Psychiatry **4**, 561–571 (1961)

4. Beck, A.T., Steer, R.A., Carbin, M.G.: Psychometric properties of the Beck Depression Inventory twenty-five years of evaluation. Clin. Psychol. Rev. **8**, 77–100 (1988)
5. Izawa, S., et al.: Salivary dehydroepiandrosterone secretion in response to acute psychosocial stress and its correlations with biological and psychological changes. Biol. Psychol. **79** (3), 294–298 (2008)
6. Ito, Y., et al.: Relationships between salivary melatonin levels, quality of sleep, and stress in young Japanese females. Int. J. Tryptophan Res. **6**(Suppl. 1), 75–85 (2013)
7. Sekiyama, A.: Interleukin-18 is involved in alteration of hipothalamic-pituitary-adrenal axis activity by stress. In: Society of Biological Psychiatry Annual Meeting, San Diego, USA (2007)
8. Kawamura, N., Shinoda, K., Ohashi, Y., Ishikawa, T., Sato, H.: Biomarker for depression, method for measuring a biomarker for depression, computer program, and recording medium. U. S. Patent, US2015126623 (2015)
9. Hagiwara, N., et al.: Validity of mind monitoring system as a mental health indicator using voice. Adv. Sci. Technol. Eng. Syst. J. **2**(3), 338–344 (2017)
10. Tokuno, S.: Pathophysiological voice analysis for diagnosis and monitoring of depression. In: Kim, Y.-K. (ed.) Understanding Depression, pp. 83–95. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-6577-4_6
11. Yang, Y., Fairbairn, C., Cohn, J.F.: Detecting depression severity from vocal prosody. IEEE Trans. Affect. Comput. **4**(2), 142–150 (2013)
12. Sheehan, D.V., et al.: The Mini-International Neuropsychiatric Interview (M.I.N.I): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J. Clin. Psychiatry **59**(Suppl. 20), 22–33 (1998)
13. Eyben, F., Wöllmer, M., Schuller, B.: Opensmile: the munich versatile and fast open-source audio feature extractor. In: Bimbo, A.D., Chang, S.F., Smeulders, A.W.M. (eds.) ACM Multimedia, pp. 1459–1462 (2010)
14. Hall, M., et al.: The WEKA data mining software: an update. ACM SIGKDD Explor. Newsl. **11**(1), 10–18 (2009)
15. Shevade, S.K., Keerthi, S.S., Bhattacharyya, C., Murthy, K.R.K.: Improvements to the SMO algorithm for SVM regression. IEEE Trans. Neural Netw. **11**, 1188–1193 (1999)
16. Zimmerman, M., Martinez, J.H., Young, D., Chelminski, I., Dalrymple, K.: Severity classification on the Hamilton depression rating scale. J. Affect. Disord. **150**(2), 384–388 (2013)